

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number  
**WO 02/098902 A2**

(51) International Patent Classification<sup>7</sup>: **C07K 1/02**  
(21) International Application Number: **PCT/IB02/02949**  
(22) International Filing Date: **31 May 2002 (31.05.2002)**  
(25) Filing Language: **English**  
(26) Publication Language: **English**

(30) Priority Data:  
0113657.1 5 June 2001 (05.06.2001) **GB**

(71) Applicant (for all designated States except US): **GENE-PROT, INC.** [US/CH]; 2, rue Pré-de-la-Fontaine, CH-1217 Meyrin 2 (CH).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **VILLAIN, Matteo** [IT/CH]; Geneprot, Inc, 2, Pre-de-la-Fontaine, CH-1217 Meyrin (CH). **GAERTNER, Hubert** [FR/FR]; 2, rue du Pre-de-la-Fontaine, CH-1217 Meyrin (CH).

(74) Agents: **HOWARD, Paul, Nicholas et al.**; Carpmals & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **IMPROVED NATIVE CHEMICAL LIGATION WITH THREE OR MORE COMPONENTS**

(57) Abstract: The invention provides a method of assembling oligopeptide intermediates in a native chemical ligation reaction that eliminates self-ligation of bi-functional intermediates. An important aspect of the invention is a bi-functional intermediate with an N-terminal cyclic thiazolidine protecting group which effectively prevents self-ligation in the chemical assembly process. The present invention is useful in methods for convergent synthesis of polypeptides and proteins and improves the efficiency of native chemical ligation reactions, particularly where three or more peptide fragments are used to assemble a polypeptide or protein product.

35- WO 02/098902 A2

**IMPROVED NATIVE CHEMICAL LIGATION  
WITH THREE OR MORE COMPONENTS**

**Field of the Invention**

The invention relates generally to methods for synthesizing polypeptides and proteins, and more particularly, to a method and intermediates for covalently assembling multiple peptide fragments into a full length polypeptide.

**BACKGROUND**

The sequencing of the human genome has created the promise and opportunity for understanding the function of all genes and proteins relevant to human biology and disease, Peltonen and McKusick, Science, 291: 1224-1229 (2001). However, several important hurdles must be overcome before this promise can be fully attained. First, the sequence signals that indicate the location of a gene in the genome and that control its expression are not well understood, so it is frequently difficult to predict the presence of a gene that is actually transcribed, e.g. Guigo et al, Genome Res., 10: 1631-1642 (2000); Rudd et al, Electrophoresis, 19:536-544 (1998). Second, although monitoring gene expression at the transcript level has become more robust with the development of microarray technology, numerous problems still exist including variability relating to probe hybridization differences and cross-reactivity, element-to-element differences within microarrays, and microarray-to-microarray differences, e.g. Audic and Claverie, Genome Res., 7: 986-995 (1997); Wittes and Friedman, J.Natl.Cancer Inst., 91: 400-401 (1999); Richmond et al, Nucleic Acids Research, 27: 3821-3835 (1999). Finally, because of the scale of human molecular biology (about a third of the estimated 30-40 thousand genes appear to give rise to multiple splice variants and most appear to encode protein products with a plethora of post-translational modifications), potentially many tens of thousands of genes and their expression products will have to be isolated and tested in order to understand their role in health and disease, Dawson and Kent, Annu.Rev.Biochem., 69: 923-960 (2000).

In regard to the issue of scale, the application of conventional recombinant methodologies for cloning, expressing, recovering, and isolating proteins are still time consuming and labor-intensive processes, so that their application in screening large numbers of different gene products for determining function has been limited. Recently, a convergent synthesis approach has been developed which may address the need for facile access to highly purified research-scale amounts of protein for functional screening, Dawson and Kent (cited above); Dawson et al, Science, 266: 776-779 (1994). In its most attractive implementation, an unprotected oligopeptide intermediate having a C-terminal thioester reacts with an N-terminal cysteine of another oligopeptide intermediate under mild aqueous conditions to form a thioester linkage which spontaneously rearranges to a natural peptide linkage, Kent et al, U.S. patent 6,184,344. The approach has

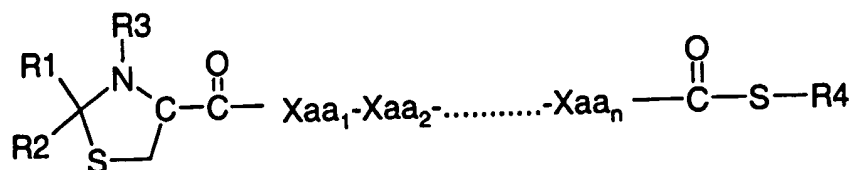
been used to assemble oligopeptides into active proteins both in solution phase, e.g. Kent et al, U.S. patent 6,184,344, and on a solid phase support, e.g. Canne et al, J.Am.Chem.Soc., 121: 8720-8727 (1999).

When the polypeptide to be synthesized by this approach exceeds 100-150 amino acids, it is necessary to join three or more fragments, as it is currently difficult to synthesize and purify oligopeptide intermediates longer than about 60 residues. In this case, the internal oligopeptide intermediates not only contain a C-terminal thioester moiety, but also an N-terminal cysteine. During the assembly process, the cysteine of such internal intermediates, if left free, will react with the C-terminal thioester of the same intermediate molecule or that of a different intermediate molecule, thereby interfering with the desired ligation reaction by the formation of an undesired cyclical peptide or concatemer of the intermediate. This problem can be circumvented by employing a protecting group for the N-terminal cysteine with the following properties: i) it must be stable to the conditions used to cleave the oligopeptide from the synthesis resin, ii) it must be removable after a native chemical ligation has been completed, and iii) preferably, removal takes place in the same ligation reaction mixture before purification, so that the ligation reaction and cysteine deprotection can be conducted in one pot.

### SUMMARY OF THE INVENTION

In view of the above, objects of the invention include, but are not limited to, providing a method for convergent synthesis of polypeptides; providing oligopeptide intermediates that can undergo native chemical ligation to form a polypeptide product in a multi-component synthesis, but that are resistant to self-ligation and concatemerization; providing a heterocyclic protecting group for terminal cysteine residues of oligopeptide intermediates of native chemical ligation reactions; providing a method of protecting thioester-modified oligopeptide intermediates from self-ligations or concatemerizations; and providing a method for native chemical ligation of successive oligopeptide intermediates in a single reaction mixture.

The invention accomplishes these and other objectives by providing a thiazolidine protecting group for N-terminal cysteines of internal thioester-modified oligopeptide intermediates of native chemical ligation reactions. Preferably, such protected intermediates are defined by the formula:



wherein:

Xaa<sub>i</sub> is an amino acid for i=1 to n;

n is an integer from 2 to 100, preferably 2 to 70, and more preferably, from 2 to 50;

$R^1$  is selected from H,  $C_{1-12}$  (preferably  $C_{1-6}$ ) alkyl groups optionally substituted with 1-3 substituents selected from halo (preferably chloro or bromo), nitro, amino, alkylamino, dialkylamino, amino oxycarbonylamino ( $H_2N-O-CH_2-CO-NH-$ ), alkylamino oxycarbonylamino ( $RHN-O-CH_2-CO-NH-$ ), dialkylamino oxycarbonylamino ( $R_2N-O-CH_2-CO-NH-$ ). Preferably,  $R^1$  is selected from the group consisting of H,  $-CH_2-NO_2$ ,  $-CH_2Cl$ ,  $-CH_2Br$ ,  $-CHCl_2$ ,  $-CHBr_2$ ;

$R^2$  is selected from H,  $C_{1-12}$  (preferably  $C_{1-6}$ ) alkyl groups optionally substituted with 1-3 substituents selected from halo (preferably chloro or bromo), nitro, amino, alkylamino, dialkylamino, amino oxycarbonylamino ( $H_2N-O-CH_2-CO-NH-$ ), alkylamino oxycarbonylamino ( $RHN-O-CH_2-CO-NH-$ ), dialkylamino oxycarbonylamino ( $R_2N-O-CH_2-CO-NH-$ ). Preferably,  $R^2$  is selected from the group consisting of H,  $H_2N-O-CH_2-CO-NH-CH_2-$ ,  $H_2N-O-CH_2-CO-NH-CHCl-$ , and  $H_2N-O-CH_2-CO-NH-CHBr-$ ;

$R^3$  is H or an amino protecting group; and

$R^4$  is alkyl having from 1 to 6 carbon atoms or alkylaryl having from 6 to 8 carbon atoms,  $-CH_2-CONH_2$ ,  $-CH_2CH_2CONH_2$ , or  $-(CH_2)_k-CO-Xaa$ , wherein k is an integer equal to 1 or 2 and Xaa is an amino acid.

Preferably, whenever  $R^1$  does not equal  $R^2$ , the invention includes both chiral variants of the compound. Preferably, whenever  $R^3$  is an amino protecting group it is selected from any amino protecting group known to those skilled in the art of peptide synthesis. Preferably, it is selected from the group consisting of tert-butyloxycarbonyl (Boc), (9H-fluoren-9-ylmethoxy)carbonyl (Fmoc), and 2-(4-nitrophenylsulfonyl) ethoxycarbonyl (Nsc).

In accordance with the method of the invention, a polypeptide is synthesized from three or more oligopeptide intermediates by the following steps: (i) forming a ligation product between an oligopeptide possessing both an N-terminal thiazolidine protected cysteine and a C-terminal thioester, and a second oligopeptide possessing an N-terminal cysteine (under conditions which permit the formation of an amide bond between the carboxylic group of the amino acid involved in the thioester linkage and the  $\alpha$ -amino group of the N-terminal cysteine of a prior ligation product) (i.e. forming a ligation product with an N-terminal thiazolidine-protected cysteine by reacting an internal oligopeptide intermediate with a prior ligation product under conditions that permit the formation of an amide bond between an  $\alpha$ -carbon of a C-terminal thioester amino acid of the internal oligopeptide intermediate and an  $\alpha$ -carbon of the N-terminal cysteine of the prior ligation product); (ii) treating the ligation product with a nucleophilic agent under acidic conditions to form a prior ligation product having a free N-terminal cysteine; and (iii) repeating steps d) and e) until the polypeptide is formed.

An important feature of the invention is the deprotection of the N-terminal cysteine by opening the N-terminal thiazolidine ring of the ligation product with a nucleophilic agent under acidic conditions after ligation and in the same reaction mixture. Such agents include, but are not limited to, O-alkylhydroxylamines and hydrazines. Preferably, thiazolidine-protected N-terminal cysteines are deprotected by treatment with an O-alkylhydroxylamine under acidic conditions. More preferably, such O-alkylhydroxylamine deprotection agent has a formula:  $H_2N-O-R$ , where R is methyl, ethyl, isopropyl, isobutyl, or  $-CH_2-COOH$ . Preferably, R is methyl. Preferably, the acidic conditions include a pH in the

range of from 2.0 to 6.0. More preferably, such pH is in the range of from 3.0 to 4.0. Preferably, thiazolidine deprotection is carried out after the amino protecting group has been removed.

The present invention is useful in methods for convergent synthesis of polypeptides and proteins and advantageously addresses limitations in these methodologies. In particular, it provides a method and materials for improving the efficiency of native chemical ligation reactions used to assemble thioester-  
5 modified oligopeptide intermediates into polypeptides or proteins, particularly in such reactions involving more than two components.

#### Brief Description of the Figures

Figure 1 illustrates the general native chemical ligation scheme related to the invention.

Figure 2 is a listing of the amino acid sequence of the polypeptide assembled from four oligopeptide intermediates described in Example 1.

Figure 3 shows analytical HPLC chromatograms of the crude peptides of Example 1: fragment 1  
15 (301; SEQ ID NO: 1), fragment 2 (302; SEQ ID NO: 2), fragment 3 (303; SEQ ID NO: 3), and fragment 4 (304; SEQ ID NO: 4).

Figure 4a shows analytical HPLC chromatograms of the ligation product of fragment 1 and fragment 2 before and after thiazolidine deprotection.

Figure 4b shows a hypermass reconstructions the raw MS data from the main HPLC peaks of the  
20 HPLC chromatograms of Fig. 4a.

Figure 5a shows analytical HPLC chromatograms of the ligation product of fragment 3 and the ligation product of fragment 1 and fragment 2 before and after thiazolidine deprotection.

Figure 5b shows a hypermass reconstructions the raw MS data from the main HPLC peaks of the  
HPLC chromatograms of Fig. 5a.

Figure 6a shows an analytical HPLC chromatogram of the final ligation product of fragments 1, 2,  
25 3, and 4.

Figure 6b shows a hypermass reconstructions the raw MS data from the main HPLC peaks of the HPLC chromatograms of Fig. 6a.

Figure 7 shows analytic HPLC chromatograms of the purified and reduced final ligation product and  
30 the crude folded material.

Figure 8 shows a hypermass reconstruction and the raw MS data from the main HPLC peak of the final folded material of Fig. 7.

Figure 9 shows the rate of cysteine deprotection for different pH conditions.

Figure 10 shows the rate of cysteine deprotection for different concentrations of the O-  
35 methylhydroxylamine deprotecting agent.

#### DEFINITIONS

The terms "polypeptide," "peptide," "peptide fragment," "oligopeptide," or "fragment" in reference to a peptide, as used herein refers to a compound made up of a single unbranched chain of amino acid  
40 residues linked by peptide bonds. The number of amino acid residues in such compounds varies widely; however, preferably, peptides or oligopeptides referred to herein usually have from 2 to 70 amino acid

residues; and more preferably, they have for 2 to 50 amino acid residues. Polypeptides and peptide fragments referred to herein usually have from a few tens of amino acid residues, e.g. 20, to up to a few hundred amino acid residues, e.g. 200, or more.

The term "protein" as used herein may be used synonymously with the term "polypeptide" or may refer to, in addition, a complex of two or more polypeptides which may be linked by bonds other than peptide bonds, for example, such polypeptides making up the protein may be linked by disulfide bonds. The term "protein" may also comprehend a family of polypeptides having identical amino acid sequences but different post-translational modifications, such as phosphorylations, acylations, glycosylations, and the like, particularly as may be added when such proteins are expressed in eukaryotic hosts.

Amino acid residues are referred to herein by their standard single-letter or three-letter notations: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, Isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine. An amino acid sequence set forth herein, such as "DKLLM," orders the amino acids from the N-terminus to the C-terminus in a left-to-right manner, unless otherwise indicated from the context.

#### DETAILED DESCRIPTION OF THE INVENTION

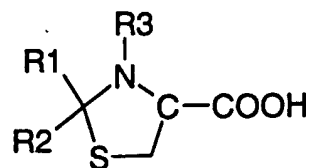
The invention relates to the assembly of oligopeptides into a polypeptide by the process of native chemical ligation, as described by Dawson et al, Science, 266: 776-779 (1994) and Kent et al, U.S. patent 6,184,344. The general approach of native chemical ligation is illustrated in Fig. 1. A first oligopeptide is provided with an N-terminal cysteine having an unoxidized sulfhydryl side chain, and a second oligopeptide is provided with a C-terminal thioester. The unoxidized sulfhydryl side chain of the N-terminal cysteine is then condensed with the C-terminal thioester to produce an intermediate oligopeptide which links the first and second oligopeptides with a  $\beta$ -aminothioester bond. The  $\beta$ -aminothioester bond of the intermediate oligopeptide then undergoes an intramolecular rearrangement to produce the oligopeptide product which links the first and second oligopeptides with an amide bond.

A problem arises in this scheme when a polypeptide is assembled from three or more fragments. In this situation, at least one fragment will have an N-terminal cysteine and a C-terminal thioester, thereby creating the possibility for self-ligation, which under conventional reaction conditions is quite significant because of the close proximity of the intra-molecular reactive moieties. In accordance with the invention, the N-terminal cysteine of an internal fragment is protected from such reactions by a cyclic thiazolidine protecting group. Preferably, such cyclic thiazolidine protecting group is a thioprolinyl group.

Oligopeptides having a C-terminal thioester (Peptide 2 of Fig. 1) may be produced as described in the following references, which are incorporated by reference: Kent et al, U.S. patent 6,184,344; Tam et al, Proc. Natl. Acad. Sci., 92: 12485-12489 (1995); Blake, Int. J. Peptide Protein Res., 17: 273 (1981); Canne et al, Tetrahedron Letters, 36: 1217-1220 (1995); Hackeng et al, Proc. Natl. Acad. Sci., 94: 7845-7850 (1997); or Hackeng et al, Proc. Natl. Acad. Sci., 96: 10068-10073 (1999); Ingenito et al, J. Am. Chem. Soc., 121: 11369-11374 (1999). Preferably, the method described by Hackeng et al (1999) is employed. Briefly, oligopeptides are synthesized on a solid phase support (described below) typically on a 0.25 mmol scale by using the in situ neutralization/HBTU activation procedure for Boc chemistry disclosed by Schnolzer et al, Int.

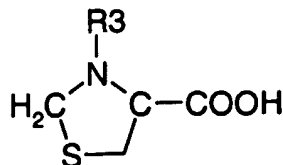
J. Peptide Protein Res., 40: 180-193 (1992), which reference is incorporated herein by reference. ( HBTU is 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate and Boc is tert-butoxycarbonyl). Each synthetic cycle consists of N<sup>α</sup>-Boc removal by a 1- to 2- minute treatment with neat TFA, a 1-minute DMF flow wash, a 10- to 20-minute coupling time with 1.0 mmol of preactivated Boc-amino acid in the presence of DIEA, and a second DMF flow wash. (TFA is trifluoroacetic acid, DMF is N,N-dimethylformamide, and DIEA is N,N-diisopropylethylamine). N<sup>α</sup>-Boc-amino acids (1.1 mmol) are preactivated for 3 minutes with 1.0 mmol of HBTU (0.5 M in DMF) in the presence of excess DIEA (3 mmol). After each coupling step, yields are determined by measuring residual free amine with a conventional quantitative ninhydrin assay, e.g. as disclosed in Sarin et al, Anal. Biochem., 117: 147-157 (1981). After coupling of Gln residues, a DCM flow wash is used before and after deprotection by using TFA, to prevent possible high-temperature (TFA/DMF)-catalyzed pyrrolidone formation. After chain assembly is completed, the oligopeptides are deprotected and cleaved from the resin by treatment with anhydrous HF for 1 hour at 0°C with 4% *p*-cresol as a scavenger. The imidazole side-chain 2,4-dinitrophenyl (dnp) protecting groups remain on the His residues because the dnp-removal procedure is incompatible with C-terminal thioester groups. However, dnp is gradually removed by thiols during the ligation reaction. After cleavage, oligopeptides are precipitated with ice-cold diethylether, dissolved in aqueous acetonitrile, and lyophilized.

Preferably, the protecting group of the invention is incorporated in the thioester oligopeptide by coupling a N-protected 2-substituted thiazolidine-4-carboxylic acid in the last synthesis cycle. Preferably, such N-protected 2-substituted thiazolidine-4-carboxylic acid has the formula:



where R<sup>1</sup> is selected from the group consisting of H, -CH<sub>2</sub>-NO<sub>2</sub>, -CH<sub>2</sub>Cl, -CH<sub>2</sub>Br, -CHCl<sub>2</sub>, -CHBr<sub>2</sub>; R<sup>2</sup> is selected from the group consisting of H, H<sub>2</sub>N-O-CH<sub>2</sub>-CO-NH-CH<sub>2</sub>-, H<sub>2</sub>N-O-CH<sub>2</sub>-CO-NH-CHCl-, and H<sub>2</sub>N-O-CH<sub>2</sub>-CO-NH-CHBr-; and where R<sup>3</sup> is an amino protecting group. Preferably, R<sup>3</sup> is Boc, Fmoc, or Nsc.

More preferably, the N-protected thiazolidine-4-carboxylic acid is an N-protected thioproline of the formula:



wherein R<sup>3</sup> is Boc or Fmoc. Most preferably, R<sup>3</sup> is Boc.

Thioester oligopeptides may be synthesized using either Fmoc or Boc chemistries. When Fmoc chemistry is employed a 3-carboxypropanesulfonamide safety catch linker is used to generate the thioester.

Thioester oligopeptides described above are preferably synthesized on a trityl-associated mercaptopropionic acid-leucine (TAMPAL) resin, made as disclosed by Hackeng et al (1999), or comparable protocol. Briefly, N<sup>α</sup>-Boc-Leu (4 mmol) is activated with 3.6 mmol of HBTU in the presence of 6 mmol of DIEA and coupled for 16 minutes to 2 mmol of p-methylbenzhydrylamine (MBHA) resin, or the equivalent. Next, 3 mmol of S-trityl mercaptopropionic acid is activated with 2.7 mmol of HBTU in the presence of 6 mmol of DIEA and coupled for 16 minutes to Leu-MBHA resin. The resulting TAMPAL resin can be used as a starting resin for polypeptide-chain assembly after removal of the trityl protecting group with two 1-minute treatments with 3.5% triisopropylsilane and 2.5% H<sub>2</sub>O in TFA. The thioester bond can be formed with any desired amino acid by using standard in situ-neutralization peptide coupling protocols for 1 hour, as disclosed in Schnolzer et al (cited above). Treatment of the final oligopeptide with anhydrous HF yields the C-terminal activated mercaptopropionic acid-leucine (MPAL) thioester oligopeptides.

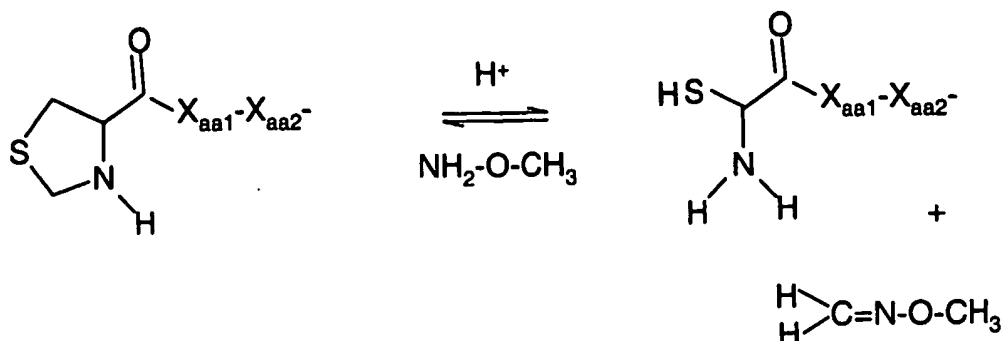
Preferably, thiazolidine-protected thioester oligopeptide intermediates are used in native chemical ligation under conditions as described by Hackeng et al (1999), or like conditions. Briefly, 0.1 M phosphate buffer (pH 8.5) containing 6 M guanidine, 2% (vol/vol) benzylmercaptan, and 2% (vol/vol) thiophenol is added to dry peptides to be ligated, to give a final peptide concentration of 1-3 mM at about pH 7, lowered because of the addition of thiols and TFA from the lyophilized peptide. Preferably, the ligation reaction is performed in a heating chamber at 37°C under continuous stirring and is periodically vortexed to equilibrate the thiol additives. The reaction may be monitored for degree of completion by MALDI-MS or HPLC and electrospray ionization MS.

After a native chemical ligation reaction is completed or stopped, the N-terminal thiazolidine ring of the product is opened by treatment with a thiazolidine deprotecting agent. A thiazolidine deprotecting agent is nucleophilic under acidic conditions, such as certain O-alkylhydroxylamines, hydrazines, or like reagents. Preferably, the N-terminal thiazolidine ring of the product is opened by treatment with O-methylhydroxylamine (0.5 M) at pH 3.5-4.5 for 2 hours at 37°C, after which a 10-fold excess of Tris-(2-carboxyethyl)-phosphine (TCEP) is added to the reaction mixture to completely reduce any oxidizing reaction constituents prior to purification of the product by conventional preparative HPLC. Preferably, fractions containing the ligation product are identified by electrospray MS, are pooled, and lyophilized. Other reducing agents that can be used in place of Tris-(2-carboxyethyl)-phosphine include β-mercaptoethanolamine, dithiotreitol, and the like.

Deprotecting the cysteine of the final product is an important feature of the invention and may be accomplished with a variety of agents that are nucleophilic under acidic conditions, as mentioned above. Opening the thiazolidine ring under acidic conditions depends on its C2 substituents (Wohr et al, J. Am. Chem. Soc., 118: 9218 (1994)). The following compounds may be used as thiazolidine deprotecting agents: O-methylhydroxylamine and other hydroxylamine derivatives, which are nucleophiles under acidic conditions, permit formation of an oxime and displacement of the equilibrium towards the unprotected form of N-terminal Cys. Hydrazine or any of its derivatives, as well as thiosemicarbazides, which are nucleophilic under acidic conditions, may also be used, but this family of reagents is more toxic than the former one and the condensation product (hydrazone, thiosemicarbazone, respectively) is less stable than the oxime. Preferably, Tris-(2-carboxyethyl)-phosphine (TCEP), or like reducing agent, is used in the deprotection reaction to rapidly and stoichiometrically reduces most peptides and sulfhydryls even under acidic conditions (Burns et al., J. Org. Chem., 56: 2648-2650, 1991). Preferably, O-methylhydroxylamine is used as the



thiazolidine deprotecting agent. O-methylhydroxylamine reacts with the masked aldehyde function in the thiazolidine ring to form an oxime, as shown below.

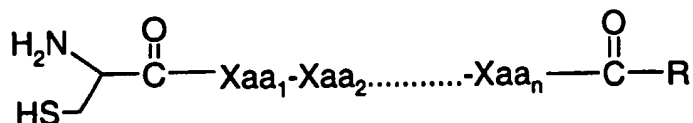


5

After the synthesis is completed and the final product purified, the final polypeptide product may be refolded by conventional techniques, e.g. Creighton, *Meth. Enzymol.*, 107: 305-329 (1984); White, *Meth. Enzymol.*, 11: 481-484 (1967); Wetlaufer, *Meth. Enzymol.*, 107: 301-304 (1984); Misawa et al, *Biopolymers*, 51: 297-307 (1999); Anfinsen, *Science*, 181: 223-230 (1973); and the like. Preferably, a final product is refolded by air oxidation by the following, or like: The reduced lyophilized product is dissolved (at about 0.1 mg/mL) in 1 M guanidine hydrochloride (or like chaotropic agent) with 100 mM Tris, 10 mM methionine, at pH 8.6. After gentle overnight stirring, the re-folded product is isolated by reverse phase HPLC with conventional protocols.

15 In a further aspect of the present invention there is provided a method of synthesizing a polypeptide by ligation of three or more oligopeptide intermediates, the method comprising the steps of:

a) providing an initial oligopeptide intermediate having an N-terminal cysteine with an  $\alpha$ -carbon;



20

wherein:

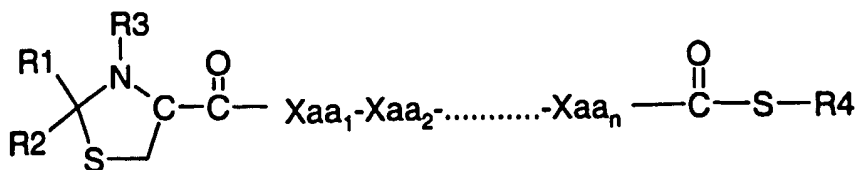
Xaai is an amino acid for  $i=1$  to  $n$ ;

$n$  is an integer from 2 to 100, preferably 2 to 70; and

R is either  $-\text{OH}$  or  $-\text{NH}_2$ ;

25

b) providing one or more internal oligopeptide intermediates each having a C-terminal amino acid thioester and an N-terminal thiazolidine-protected cysteine, the oligopeptide intermediate initially having the formula:



wherein:

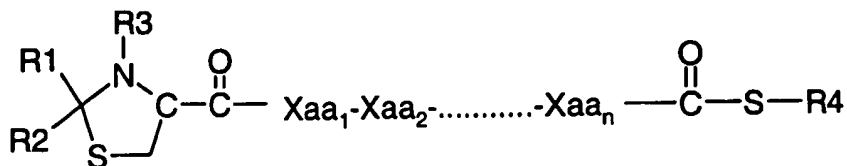
Xaa<sub>i</sub>, n, and R<sup>4</sup> are defined as above;

n is an integer from 2 to 100, preferably 2 to 70;

R<sup>1</sup> and R<sup>2</sup> are hydrogen;

5 R<sup>3</sup> is hydrogen or an amino protecting group selected from the group consisting of Boc and Fmoc;

c) providing a terminal oligopeptide intermediate having a C-terminal amino acid thioester and, whenever such terminal oligopeptide intermediate has an N-terminal cysteine, an N-terminal thiazolidine-protected cysteine;



10

wherein Xaa<sub>i</sub>, n, R<sup>1</sup>, R<sup>2</sup>, R<sup>3</sup>, and R<sup>4</sup> are defined as above (in a preferred embodiment, R<sup>1</sup>=R<sup>2</sup>=R<sup>3</sup>=H although it will be appreciated that according to the synthesis procedure R<sup>3</sup> will be a protecting group such as Boc or Fmoc, but for the cleavage from the resin, which means R<sup>4</sup> no longer attached to the linker of the solid phase support, the secondary amine will be deprotected);

15

d) forming a ligation product with an N-terminal thiazolidine-protected cysteine by reacting an internal oligopeptide intermediate with the initial oligopeptide intermediate or a prior ligation product having an N-terminal cysteine with a free α-amine under conditions that permit the formation of an amide bond between the carboxylic group of the C-terminal thioester amino acid of the internal oligopeptide intermediate and the free α-amine of the N-terminal cysteine of the prior ligation product or the initial oligopeptide intermediate;

20

e) treating the ligation product with an O-alkylhydroxylamine under acidic conditions in the range of from pH 2.0 to 6.0 to form a prior ligation product having a free N-terminal cysteine;

f) repeating steps d) and e) until all of the internal oligopeptide intermediates are ligated; and

25

g) forming the polypeptide by reacting the terminal oligopeptide intermediate with the prior ligation product under conditions that permit the formation of an amide bond between the α-amine of the N-terminal cysteine of the prior ligation product and the carboxylic group of the amino acid involved in the thioester linkage of the N-terminal oligopeptide intermediate.

30

### Example 1

In this example, a polypeptide having the sequence of Fig. 2 (SEQ ID NO: 5) was synthesized with the method and materials of the invention. The full length polypeptide was assembled from the previously synthesized oligopeptide intermediates listed below (the superscripted numbers indicate the position of the fragments in the sequence of Fig. 2). Fragment 1 was initially coupled to fragment 2 to give a first product,

35

then after preparative HPLC purification, the first product was coupled to fragment 3 to give a second product. Again, after preparative HPLC purification, the second product was coupled to fragment 4 to give the desired polypeptide, which was purified and refolded.

Fragment	SEQ ID NO	Sequence of Oligopeptide Intermediate
1	1	C <sup>139</sup> AEKSDYIRK INELMPKYAP KAASARTDL <sup>139</sup>
2	2	C <sup>82</sup> EKLKKDSQ ICELYDKQI DLSTVDLKKL RVKELKKILD DWGETCKG <sup>139</sup>
3	3	C <sup>10</sup> REARGKENR LCYYIGATDD AATKIINEVS KPLAHHIPVE KI <sup>11</sup>
4	4	L <sup>1</sup> RPGDCEVCI SYLGRFYQDL KDRDVTFSFA TIENELIKF <sup>19</sup>

5

Thioester formation. Fragments 2, 3, 4 were synthesized on a thioester generating resin. For this purpose S-acetylthioglycolic acid pentafluorophenylester was coupled to a Leu-PAM resin under conditions essentially as described by Hackeng et al (1999). The resulting resin was used as a starting resin for peptide chain elongation on a 0.2 mmol scale after removal of the acetyl protecting group with a 30 min treatment with 2M mercaptoethanol, 2M piperidine in DMF. The thioester was formed with Boc-Gly-OH for synthesis of fragment 2, Boc-Ile-OH for fragment 3 and Boc-Phe-OH for fragment 4 using the standard in situ neutralization coupling protocol for 1 hour, Schnölzer et al (cited above), with a 4-fold molar excess of aa over the sulfhydryl group. The N<sup>α</sup> of the N-terminal Cys residues of fragments 2 and 3 were protected in accordance with the invention by coupling a Boc-thioproline (Boc-SPr, i.e. Boc-L-thioproline, ) to the terminus of the respective chains instead of a Cys having conventional N<sup>α</sup> or S<sup>β</sup> protection, e.g. Brik et al, J. Org. Chem., 65: 3829-3835 (2000)

15

Peptide synthesis. Solid-phase synthesis was performed on a custom-modified 433A peptide synthesizer from Applied Biosystems, using in situ neutralization/2-(1H-benzotriazol-1-yl)-1,1,1,3,3-tetramethyluronium hexafluoro-phosphate (HBTU) activation protocols for stepwise Boc chemistry chain elongation, as described by Schnolzer et al, Int. J. Peptide Protein Res., 40: 180-193 (1992). Each synthetic cycle consisted of N<sup>α</sup>-Boc -removal by a 1 to 2 min treatment with neat TFA, a 1-min DMF flow wash, a 10-min coupling time with 2.0 mmol of preactivated Boc-amino acid in the presence of excess DIEA and a second DMF flow wash. N<sup>α</sup>-Boc-amino acids (2 mmol) were preactivated for 3min with 1.8mmol HBTU (0.5M in DMF) in the presence of excess DIEA (6mmol). After coupling of Gln residues, a dichloromethane flow wash was used before and after deprotection using TFA, to prevent possible high temperature (TFA/DMF)-catalyzed pyrrolidone carboxylic acid formation. Side-chain protected amino acids were Boc-Arg(p-toluenesulfonyl)-OH, Boc-Asn(xanthyl)-OH, Boc-Asp(O-cyclohexyl)-OH, Boc-Cys(4-methylbenzyl)-OH, Boc-Glu(O-cyclohexyl)-OH, Boc-His(dinitrophenylbenzyl)-OH, Boc-Lys(2-Cl-Z)-OH, Boc-Ser(benzyl)-OH, Boc-Thr(benzyl)-OH, Boc-Trp(formyl)-OH and Boc-Tyr(2-Br-Z)-OH (Orpegen Pharma, Heidelberg, Germany). Other amino acids were used without side chain protection. C-terminal Fragment 1 was synthesized on Boc-Leu-O-CH<sub>2</sub>-Pam resin (0.71mmol/g of loaded resin), while for Fragments 2 ,3 and 4 machine-assisted synthesis was started on the Boc-Xaa-S-CH<sub>2</sub>-CO-Leu-Pam resin . Figure 3 shows analytical HPLC chromatograms of the crude peptides of fragment 1 (301), fragment 2 (302), fragment 3 (303), and fragment 4 (304).

30

After chain assembly was completed, the peptides were deprotected and cleaved from the resin by treatment with anhydrous hydrogen fluoride for 1hr at 0°C with 5% p-cresol as a scavenger. In all cases except Fragment 1, the imidazole side chain 2,4-dinitrophenyl (DNP) protecting groups remained on His residues because the DNP-removal procedure is incompatible with C-terminal thioester groups. However DNP is gradually removed by thiols during the ligation reaction, yielding unprotected His. After cleavage, both peptides were precipitated with ice-cold diethylether, dissolved in aqueous acetonitrile containing 0.1% TFA and lyophilized. The peptides were purified by RP-HPLC with a C18 column from Waters by using linear gradients of buffer B (acetonitrile/0.1% trifluoroacetic acid) in buffer A (H<sub>2</sub>O/0.1% trifluoroacetic acid) and UV detection at 214nm. Collected fractions were analyzed by electrospray mass spectrometry (ESMS) using an Esquire instrument (Brücker, Bremen, Germany) and by analytical HPLC. Fractions containing the right product were pooled and freeze-dried.

Native chemical ligations. As described more fully below, the ligation of unprotected fragments was performed as follows: the dry peptides were dissolved in equimolar amounts in 6M guanidine hydrochloride (GuHCl), 0.2M phosphate, pH 7.5 in order to get a final peptide concentration of 1-5 mM at a pH around 7, and 2% thiophenol was added. Usually, the reaction was carried out overnight and was monitored by HPLC and electrospray mass spectrometry. The ligation product was subsequently treated to hydrolyze any remaining thioester and to remove protecting groups still present. For this purpose and in order to remove the formyl group of Trp, 20% 2-mercaptoethanol was added and the pH shifted to 9.0 by addition of hydrazine and the solution incubated for 1h at 37°C. The reaction mixture was then acidified to pH 3.5 with 6M HCl and 2M O-methylhydroxylamine in 6M guanidinehydrochloride, pH 3.5 added to get a final 0.5M concentration. A 2h incubation at 37°C is required for complete opening of the N-terminal thiazolidine ring. A 10-fold excess of Tris(2-carboxyethyl)phosphine over the fragment was added and the material purified by preparative HPLC after 15 min incubation. Fractions containing the polypeptide chain were identified by ESMS, pooled and lyophilized.

The ligation of SP<sup>82</sup>-(82-129)-αCOSR (fragment 2) and Cys<sup>130</sup>(130-158)-COOH (fragment 1) peptides was performed at pH 7.0 in 6 M GuHCl. The concentration of each reactant was 5mM, and 2% thiophenol were added to create a reducing environment and facilitate the ligation reaction. An almost quantitative ligation reaction is observed after overnight stirring at 37°C. ESMS confirmed the molecular mass for the SP<sup>82</sup>-(82-158)-COOH polypeptide chain of 8960.8Da, in good agreement with the calculated average isotopic mass of 8960.5Da. At this point in the reaction, 2-mercaptoethanol was added to a 20%(v/v) final concentration and hydrazine added to shift the pH to 9.0, for the removal of the formyl group of Trp<sup>123</sup>. After a 1h incubation at 37°C, the medium was acidified, CH<sub>3</sub>-O-NH<sub>2</sub>.HCl was added to the solution to give a 0.5M final concentration, and the pH adjusted to 3.5 in order to open the thiazolidine ring of SP<sup>82</sup>. After 2h incubation at 37°C, ESMS confirmed the removal of the two protecting groups (Table 1), corresponding to the loss of 28 and 12 Da, respectively. The reaction mixture was subsequently treated with a 10-fold excess of Tris(2-carboxyethylphosphine) over the peptide and after 15min, the ligation product was purified using the preparative HPLC (C4, 20-60% CH<sub>3</sub>CN, 0.5% per min and a flow rate of 20mL/min), lyophilised and stored at -20°C. Figure 4a shows analytical HPLC chromatograms of the ligation product of fragment 1 and fragment 2 where curve (401) is the chromatogram of the product before thiazolidine deprotection and curve (402) is the chromatogram of the product after thiazolidine deprotection. Figure 4b shows a hypermass reconstructions the raw MS data from the main HPLC peaks of the HPLC chromatograms of Fig.

4a, wherein curve (403) corresponds to the peak of curve (401) (closed thiazolidine ring) and curve (404) corresponds to the peak of curve (402) (open thiazolidine ring).

The same procedure was repeated for ligation 2 with some slight modifications. Since ligation reaction 2 involves an Ile-Cys ligation site, which is one of the less favourable ones according to a comparative study (Hackeng et al, P.N.A.S., 96, 10068-10073, 1999), the ligation reaction was carried over 48h. According to the HPLC profile, only a 60% ligation yield could be obtained after 48h incubation. Figure 5a shows analytical HPLC chromatograms of the ligation product of the above ligation product and fragment 3 where curve (501) is the chromatogram of the product before thiazolidine deprotection and curve (502) is the chromatogram of the product after thiazolidine deprotection. Figure 5b shows a hypermass reconstruction the raw MS data from the main HPLC peaks of the HPLC chromatograms of Fig. 5a, wherein curve (503) corresponds to the peak of curve (501) (closed thiazolidine ring) and curve (504) corresponds to the peak of curve (502) (open thiazolidine ring).

Ligation 3 didn't require any further work-up of the full length peptide: the ligation medium was immediately purified after acidification and TCEP treatment. Figure 6a shows an analytical HPLC chromatogram of the final ligation product of fragments 1, 2, 3, and 4. Figure 6b shows a hypermass reconstruction of the raw MS data from the main peak of the chromatogram of Fig. 6a. Figure 7 shows analytic HPLC chromatograms of the purified and reduced final ligation product (710) and the crude folded material (711). Figure 8 shows a hypermass reconstruction (801) and the raw MS data (802) from the main HPLC peak of the final folded material of Fig. 7 curve (711).

Polypeptide Folding. The full length peptide was refolded by dissolving the reduced lyophilized protein (about 0.1 mg/mL) in 1M GuHCl, 100mM Tris, 0.5mM reduced glutathione, 0.1mM oxidized glutathione, 10mM methionine, pH 8.6 After gentle stirring overnight, the protein solution was purified by RP-HPLC as described above. The masses of the reduced and refolded proteins agreed well with their calculated average isotopic mass as summarized in Table 1.

25

**Table I**  
ESMS Characterization of Oligopeptide Intermediates in the  
Synthesis of Full Length Product, Gln<sup>1</sup>-Leu<sup>161</sup>

Fragment	Calculated Mass (Daltons, average isotope)	Measured Mass (Daltons, average isotope)
(Cys <sup>130</sup> -Leu <sup>158</sup> )-COOH	3283.2	3283.2
(SPR <sup>82</sup> -Gly <sup>124</sup> , Trp <sup>124</sup> (CHO))-αCOSR	5882.0	5882.3
(SPR <sup>40</sup> -Ile <sup>81</sup> , His <sup>74,75</sup> (DNP))-αCOSR	5255.95	5256.1
(Leu <sup>1</sup> -Phe <sup>39</sup> )-αCOSR	4740.49	4739.5
Ligation 1 product Cys <sup>82</sup> -(82-158)-COOH	8920.54	8920.0

Ligation 2 product Cys <sup>40</sup> -(40-158)-COOH	13626.99	13627.8
Ligation 3 product Leu <sup>1</sup> -(1-158)-COOH	18162.19	18163.5
Refolded Leu <sup>1</sup> (1-158)-COOH	18154.19	18154.3

### Example 2

A polypeptide with the sequence CYAKYAKL (SEQ ID NO: 6) was synthesized using in situ neutralization/2-(1H-benzotriazol-1-yl)-1,1,1,3,3-tetramethyluronium hexafluoro-phosphate (HBTU) activation protocol for stepwise Boc chemistry as previously described (see example 1). The N-terminal cysteine was introduced as Boc-Thioproline, with standard HBTU activation. After Boc removal the peptide was cleaved from the resin by treatment with hydrogen fluoride for 1 hr at 0° with 5% p-cresol as scavenger. The purified final product presented the expected mass of 970.18, consistent with a stable thioproline N-terminal cysteine, confirming the stability of the thioproline ring in the cleavage conditions.

A series of experiments to determine the optimal conditions for thiazolidine ring opening were conducted at different pH and different O-methylhydroxylamine concentrations. The extent of ring opening was calculated by analytical RP-HPLC. The ring opening induced a net shift in elution time that allows the complete resolution of the two peptides. MS analysis of the products confirmed a loss of 12 amu, consistent with the formation of a free cysteine at the N-terminus of the peptide.

Optimal pH assay: Aliquots of 100 µg of peptide were dissolved in a solution 6 M Guanidine hydrochloride, 0.5 M O-methylhydroxylamine, corrected either at pH 4.5, or pH4.0, or pH 3.5, or pH 3.0 and pH 2.5. Samples of the reaction media were analyzed at time 30, 90, 130, and 250 minutes. Peaks corresponding to SPrYAKYAKL (SEQ ID NO: 7) and CYAKYAKL were integrated and the extent of ring opening calculated base on the % of the CYAKYAKL peak. As evident from Figure 9, the optimal conditions for conversion of thioproline to cysteine are at pH 3.5, or higher. At pH lower then 3.5 the reaction becomes slower, which is possibly related to the pKa of the nitrogen of the O-methylhydroxylamine.

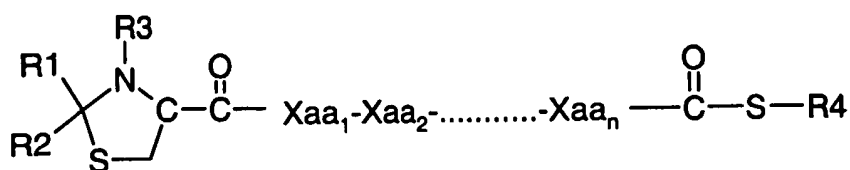
O-methylhydroxylamine optimal concentration assay. Aliquots of 100 µg of peptide were dissolved in a solution 6 M Guanidine hydrochloride with either 0.5 M O-methylhydroxylamine, or 0.25 M O-methylhydroxylamine, or 0.125 M O-methylhydroxylamine or 0.0625 M O-methylhydroxylamine, all adjusted to pH 3.5. Samples of the reaction media were analyzed at the following times: 30, 90, 130, 250 minutes. Peaks corresponding to SPrYAKYAKL and CYAKYAKL were integrated and the extent of ring opening calculated base on the % of the CYAKYAKL peak. As evident from Figure 10, a combination of high O-methylhydroxylamine concentration and pH of about 3.5 ensure the complete conversion of the N-terminal thioprolinyl to cysteinyl.

The descriptions of the foregoing embodiments of the invention have been presented for purpose of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise forms disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention to

thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

## WE CLAIM:

1. A method of synthesizing a polypeptide by ligation of three or more oligopeptide intermediates, the method comprising the steps of:
  - a) forming a ligation product between an oligopeptide possessing both an N-terminal thiazolidine protected cysteine and a C-terminal thioester, and a second oligopeptide possessing an N-terminal cysteine;
  - b) treating the ligation product with a nucleophilic agent under acidic conditions to form a prior ligation product having a free N-terminal cysteine; and
  - c) repeating steps a) and b) until the polypeptide is formed.
2. The method of claim 1 further including the step of purifying said prior ligation product prior to said step of forming.
3. The method of claim 2 wherein said internal oligopeptide intermediate is defined by the formula:

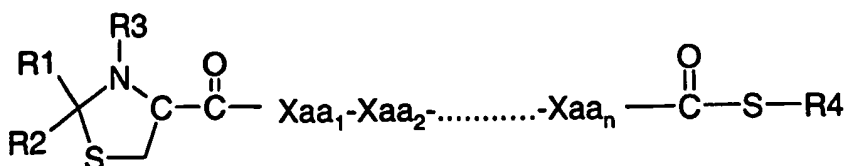


wherein:

- Xaa<sub>i</sub> is an amino acid for i=1 to n;
- n is an integer from 2 to 100;
- R<sup>1</sup> is selected from H, C<sub>1-12</sub> (preferably C<sub>1-6</sub>) alkyl groups optionally substituted with 1-3 substituents selected from halo (preferably chloro or bromo), nitro, amino, alkylamino, dialkylamino, amino oxycarbonylamino (H<sub>2</sub>N-O-CH<sub>2</sub>-CO-NH-), alkylamino oxycarbonylamino (RHN-O-CH<sub>2</sub>-CO-NH-), dialkylamino oxycarbonylamino (R<sub>2</sub>N-O-CH<sub>2</sub>-CO-NH-);
- R<sup>2</sup> is selected from H, C<sub>1-12</sub> (preferably C<sub>1-6</sub>) alkyl groups optionally substituted with 1-3 substituents selected from halo (preferably chloro or bromo), nitro, amino, alkylamino, dialkylamino, amino oxycarbonylamino (H<sub>2</sub>N-O-CH<sub>2</sub>-CO-NH-), alkylamino oxycarbonylamino (RHN-O-CH<sub>2</sub>-CO-NH-), dialkylamino oxycarbonylamino (R<sub>2</sub>N-O-CH<sub>2</sub>-CO-NH-);
- R<sup>3</sup> is H or an amino protecting group; and
- R<sup>4</sup> is alkyl having from 1 to 6 carbon atoms or alkylaryl having from 6 to 8 carbon atoms, -CH<sub>2</sub>-CONH<sub>2</sub>, -CH<sub>2</sub>CH<sub>2</sub>CONH<sub>2</sub>, or -(CH<sub>2</sub>)<sub>k</sub>-CO-Xaa, wherein k is an integer equal to 1 or 2 and Xaa is an amino acid.



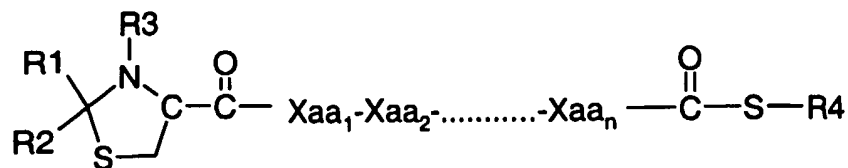
4. The method of claim 3 wherein  $R^1$  is selected from the group consisting of H,  $-\text{CH}_2\text{NO}_2$ ,  $-\text{CH}_2\text{Cl}$ ,  $-\text{CH}_2\text{Br}$ ,  $-\text{CHCl}_2$ ,  $-\text{CHBr}_2$ .
5. The method of claim 3 or 4 wherein  $R^2$  is selected from the group consisting of H,  $\text{H}_2\text{N}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-\text{CH}_2-$ ,  $\text{H}_2\text{N}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-\text{CHCl}-$ , and  $\text{H}_2\text{N}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-\text{CHBr}-$ .
6. The method of claim 3, 4 or 5 wherein said nucleophilic agent is an O-alkylhydroxylamine of the formula  $\text{H}_2\text{N}-\text{O}-\text{R}$ , where R is methyl, ethyl, isopropyl, isobutyl, or  $-\text{CH}_2-\text{COOH}$ .
7. The method of claim 6 wherein said acidic conditions are in the range of pH 2.0 to pH 6.0.
8. The method of claim 7 wherein  $R^1$  and  $R^2$  are both hydrogen and wherein  $R^3$  is Fmoc, Boc, or Nsc, whenever  $R^3$  is said amino protecting group.
9. The method of claim 8 wherein said O-alkylhydroxylamine is O-methylhydroxylamine.
10. A composition of matter comprising an N-terminal thiazolidine-protected oligopeptide intermediate of the formula:



wherein:

- $\text{Xaa}_i$  is an amino acid for  $i=1$  to  $n$ ;
- $n$  is an integer from 2 to 100;
- $R^1$  is selected from H,  $\text{C}_{1-12}$  (preferably  $\text{C}_{1-6}$ ) alkyl groups optionally substituted with 1-3 substituents selected from halo (preferably chloro or bromo), nitro, amino, alkylamino, dialkylamino, amino oxycarbonylamino ( $\text{H}_2\text{N}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-$ ), alkylamino oxycarbonylamino ( $\text{RHN}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-$ ), dialkylamino oxycarbonylamino ( $\text{R}_2\text{N}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-$ );
- $R^2$  is selected from H,  $\text{C}_{1-12}$  (preferably  $\text{C}_{1-6}$ ) alkyl groups optionally substituted with 1-3 substituents selected from halo (preferably chloro or bromo), nitro, amino, alkylamino, dialkylamino, amino oxycarbonylamino ( $\text{H}_2\text{N}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-$ ), alkylamino oxycarbonylamino ( $\text{RHN}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-$ ), dialkylamino oxycarbonylamino ( $\text{R}_2\text{N}-\text{O}-\text{CH}_2-\text{CO}-\text{NH}-$ );
- $R^3$  is H or an amino protecting group; and
- $R^4$  is alkyl having from 1 to 6 carbon atoms or alkylaryl having from 6 to 8 carbon atoms,  $-\text{CH}_2-\text{CONH}_2$ ,  $-\text{CH}_2\text{CH}_2\text{CONH}_2$ , or  $-(\text{CH}_2)_k-\text{CO}-\text{Xaa}$ , wherein  $k$  is an integer equal to 1 or 2 and Xaa is an amino acid.

11. The composition of claim 10 wherein  $R^1$  is selected from the group consisting of H,  $-\text{CH}_2\text{-NO}_2$ ,  $-\text{CH}_2\text{Cl}$ ,  $-\text{CH}_2\text{Br}$ ,  $-\text{CHCl}_2$ ,  $-\text{CHBr}_2$ .
12. The composition of claim 10 or 11 wherein  $R^2$  is selected from the group consisting of H,  $\text{H}_2\text{N-O-CH}_2\text{-CO-NH-CH}_2\text{-}$ ,  $\text{H}_2\text{N-O-CH}_2\text{-CO-NH-CHCl-}$ , and  $\text{H}_2\text{N-O-CH}_2\text{-CO-NH-CHBr-}$ .
13. The composition of claim 10, 11 or 12 wherein  $R^1$  and  $R^2$  are both hydrogen and wherein  $R^3$  is Fmoc, Boc, or Nsc, whenever  $R^3$  is said amino protecting group.
14. A method of converting an N-terminal thiazolidine-protected cysteine of an oligopeptide to a free N-terminal cysteine, the method comprising the step of treating the N-terminal thiazolidine-protected cysteine with a nucleophilic agent under acidic conditions.
15. The method of claim 14 wherein said nucleophilic agent is selected from the group consisting of O-alkylhydroxylamines, hydrazines, and TCEP.
16. The method of claim 15 wherein said nucleophilic agent is an O-alkylhydroxylamine of the formula:  $\text{H}_2\text{N-O-R}$ , where R is methyl, ethyl, isopropyl, isobutyl, or  $-\text{CH}_2\text{-COOH}$ .
17. The method of claim 16 wherein said acidic conditions include a pH in the range of 2.0 to 6.0.
18. A method of synthesizing a polypeptide by ligation of three or more oligopeptide intermediates, the method comprising the steps of:
- a) providing an initial oligopeptide intermediate having an N-terminal cysteine with an  $\alpha$ -carbon;
- $$\begin{array}{c} \text{H}_2\text{N} \\ \diagup \\ \text{C} \\ \diagdown \\ \text{HS} \end{array} - \overset{\text{O}}{\parallel}{\text{C}} - \text{Xaa}_1 - \text{Xaa}_2 \dots \dots \text{Xaa}_n - \overset{\text{O}}{\parallel}{\text{C}} - \text{R}$$
- wherein:
- Xaa<sub>i</sub> is an amino acid for  $i=1$  to  $n$ ;
- $n$  is an integer from 2 to 100; and
- R is either  $-\text{OH}$  or  $-\text{NH}_2$ ;
- b) providing one or more internal oligopeptide intermediates each having a C-terminal amino acid thioester and an N-terminal thiazolidine-protected cysteine, the oligopeptide intermediate initially having the formula:



wherein:

Xaa<sub>i</sub>, n, and R<sup>4</sup> are defined as above;

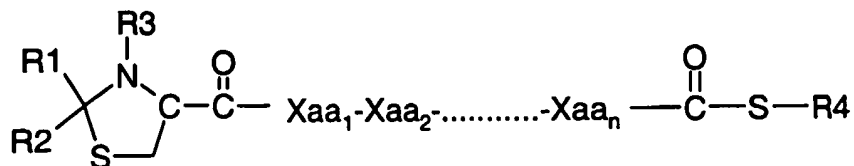
5 n is an integer from 2 to 100;

R<sup>1</sup> and R<sup>2</sup> are hydrogen;

R<sup>3</sup> is hydrogen or an amino protecting group selected from the group consisting of Boc and Fmoc;

c) providing a terminal oligopeptide intermediate having a C-terminal amino acid thioester and, whenever such terminal oligopeptide intermediate has an N-terminal cysteine, an N-terminal thiazolidine-protected cysteine;

10



wherein Xaa<sub>i</sub>, n, R<sup>1</sup>, R<sup>2</sup>, R<sup>3</sup>, and R<sup>4</sup> are defined as above;

d) forming a ligation product with an N-terminal thiazolidine-protected cysteine by reacting an internal oligopeptide intermediate with the initial oligopeptide intermediate or a prior ligation product having an N-terminal cysteine with a free α-amine under conditions that permit the formation of an amide bond between the carboxylic group of the C-terminal thioester amino acid of the internal oligopeptide intermediate and the free α-amine of the N-terminal cysteine of the prior ligation product or the initial oligopeptide intermediate;

15

20

e) treating the ligation product with an O-alkylhydroxylamine under acidic conditions in the range of from pH 2.0 to 6.0 to form a prior ligation product having a free N-terminal cysteine;

f) repeating steps d) and e) until all of the internal oligopeptide intermediates are ligated; and

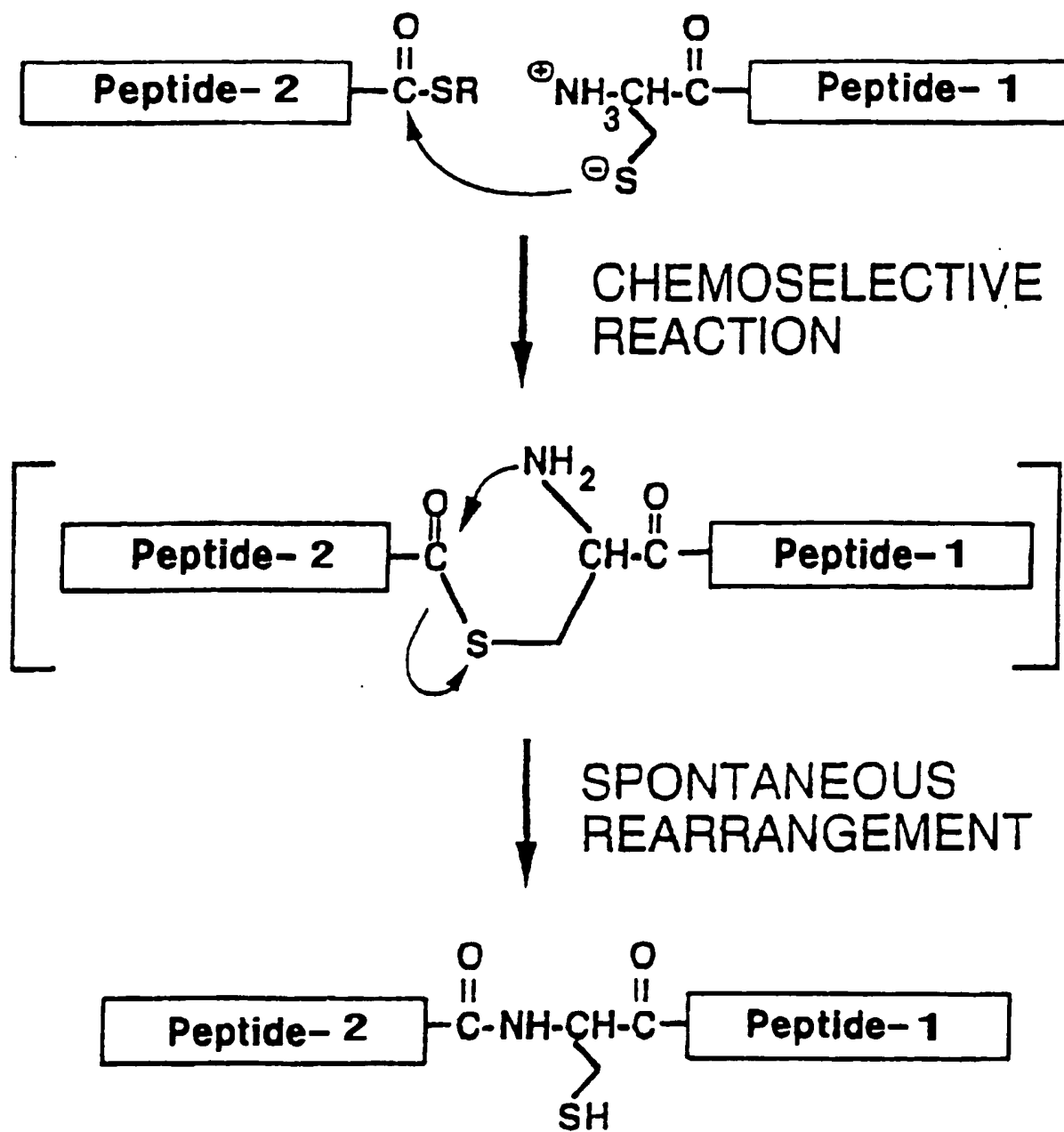
g) forming the polypeptide by reacting the the terminal oligopeptide intermediate with the prior ligation product under conditions that permit the formation of an amide bond between the α-amine of the N-terminal cysteine of the prior ligation product and the carboxylic group of the amino acid involved in the thioester linkage of the N-terminal oligopeptide intermediate.

25

30

1/13

FIG. 1



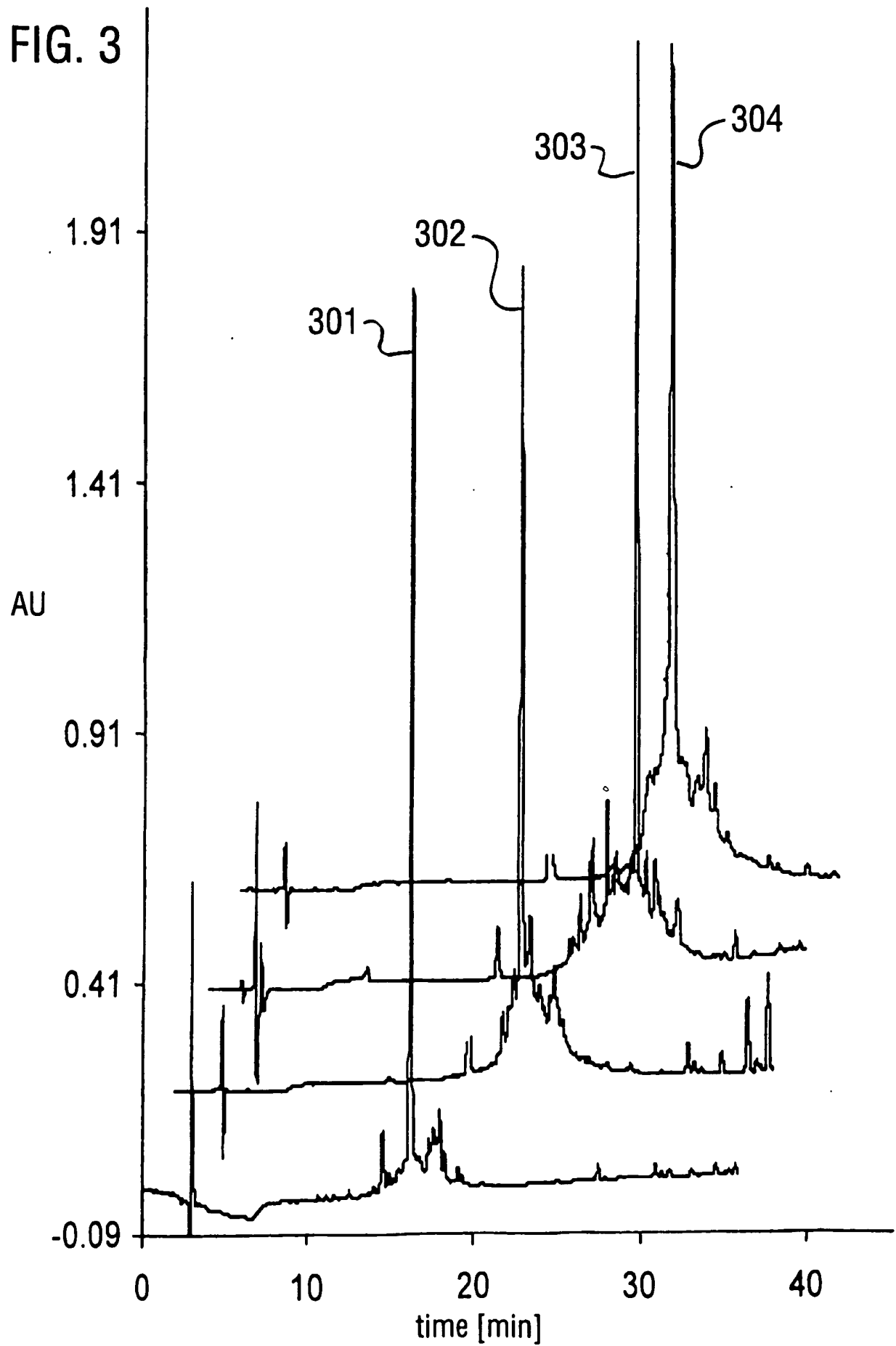
2/13

## FIG. 2

Leu	Arg	Pro	Gly	Asp	Cys	Glu	Val	Cys	Ile	Ser	Tyr	Leu	Gly	Arg	Phe			
				5					10					15				
Tyr	Gln	Asp	Leu	Lys	Asp	Arg	Asp	Val	Thr	Phe	Ser	Pro	Ala	Thr	Ile			
			20					25					30					
Glu	Asn	Glu	Leu	Ile	Lys	Phe	Cys	Arg	Glu	Ala	Arg	Gly	Lys	Glu	Asn			
			35				40					45						
Arg	Leu	Cys	Tyr	Tyr	Ile	Gly	Ala	Thr	Asp	Asp	Ala	Ala	Thr	Lys	Ile			
			50			55					60							
Ile	Asn	Gly	Val	Ser	Lys	Pro	Leu	Ala	His	His	Ile	Pro	Val	Gly	Lys			
					70					75					80			
Ile	Cys	Glu	Lys	Leu	Lys	Lys	Leu	Asp	Ser	Gln	Ile	Cys	Glu	Leu	Lys			
				85				90						95				
Tyr	Asp	Lys	Gln	Ile	Asp	Leu	Ser	Thr	Val	Asp	Leu	Lys	Lys	Leu	Arg			
			100					105					110					
Val	Lys	Glu	Leu	Cys	Lys	Ile	Leu	Asp	Asp	Trp	Gly	Glu	Thr	Cys	Lys			
			115				120					125						
Gly	Cys	Ala	Glu	Lys	Ser	Asp	Tyr	Ile	Arg	Lys	Ile	Asn	Glu	Leu	Met			
			130			135					140							
Pro	Lys	Tyr	Ala	Pro	Lys	Ala	Ala	Ser	Ala	Arg	Thr	Asp	Leu					
					150					155								

3/13

FIG. 3



4/13

FIG. 4a

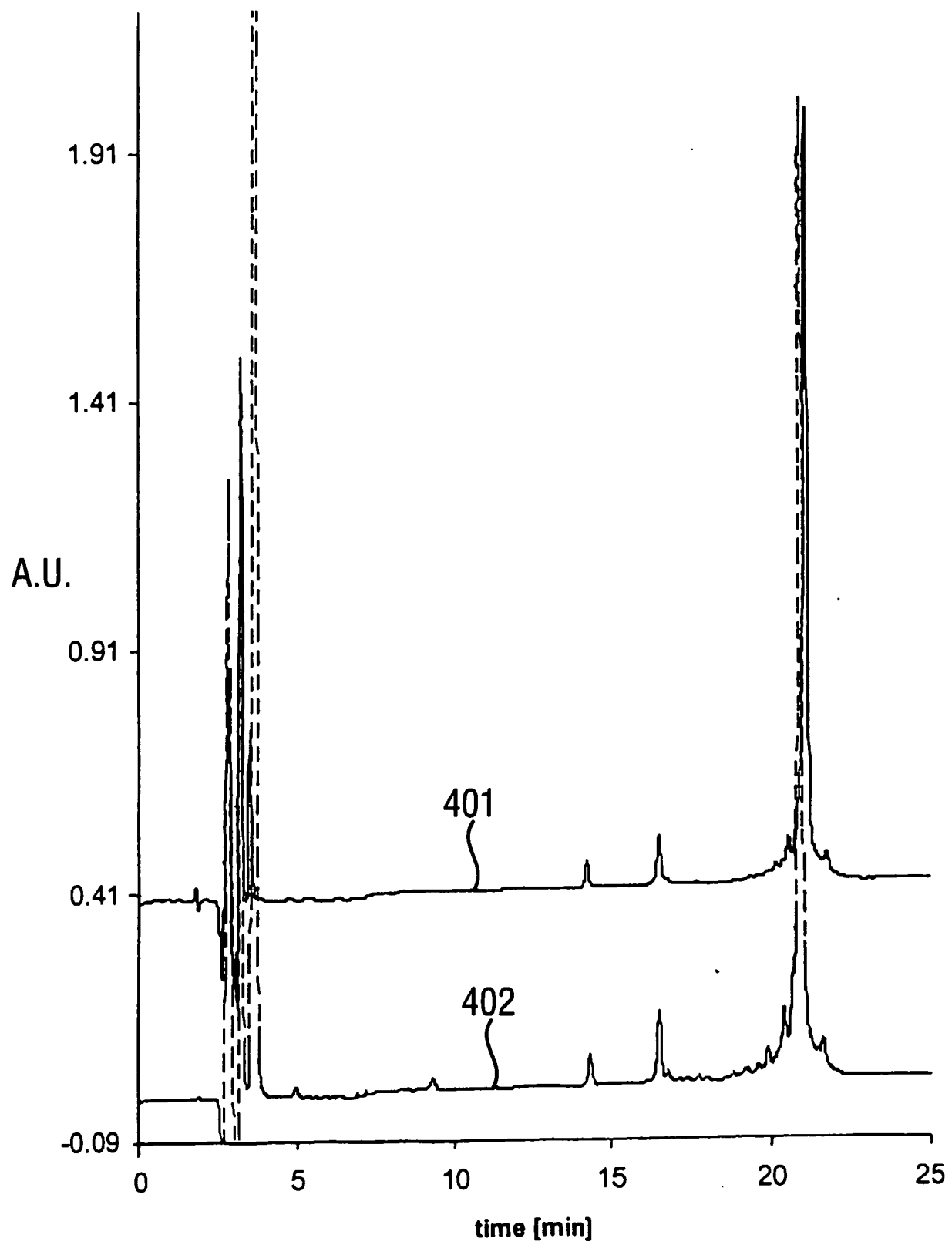
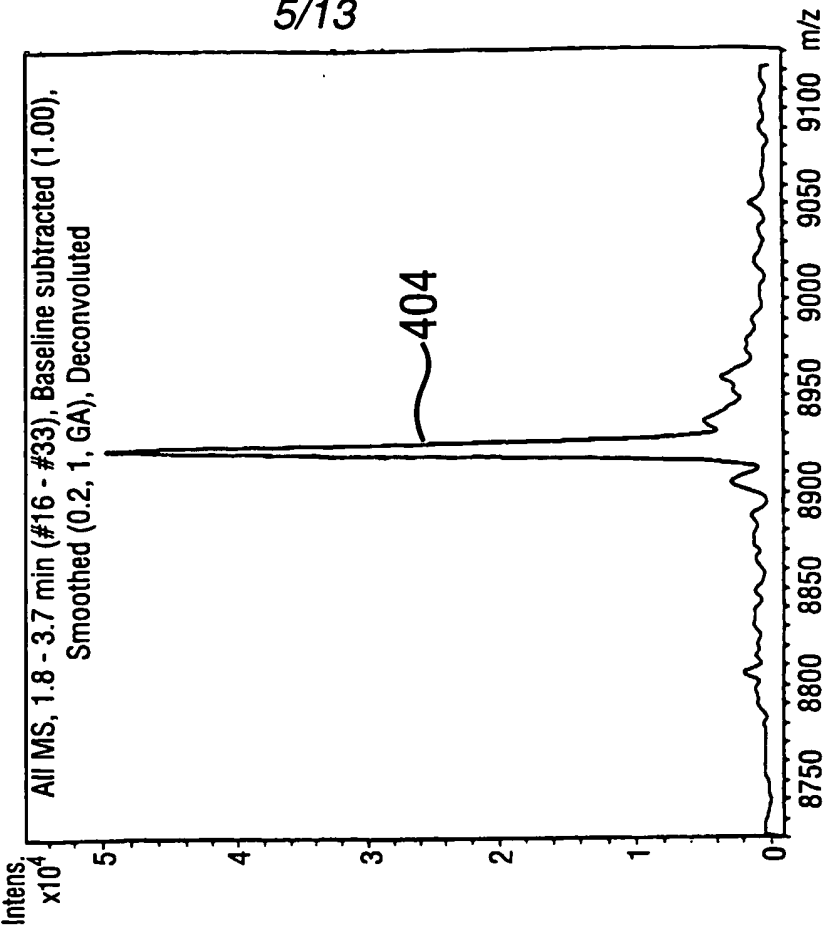
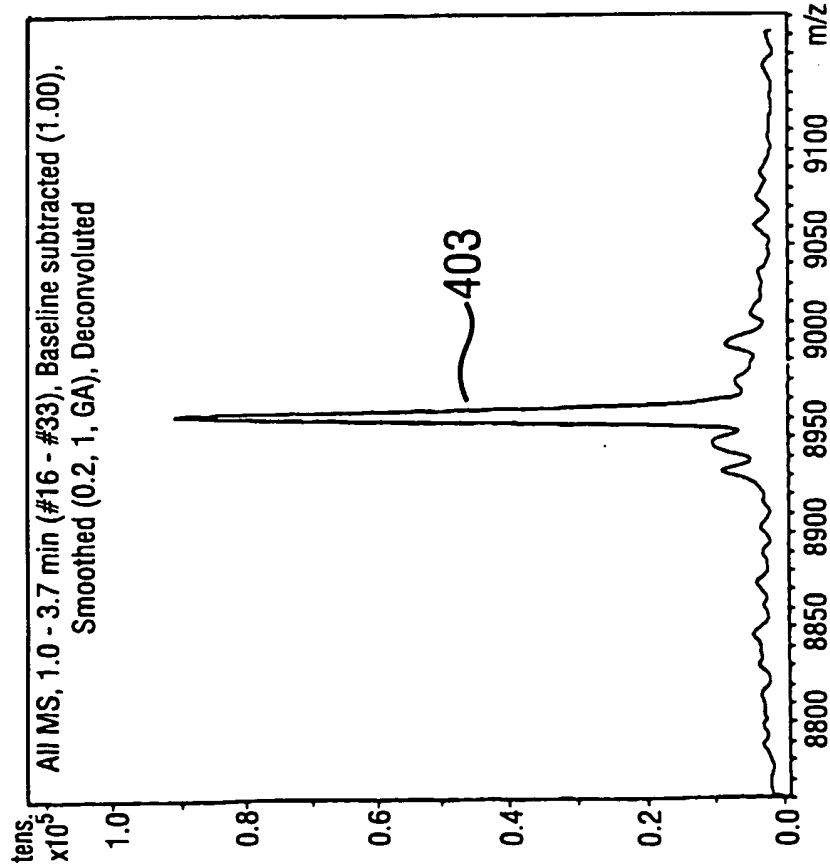


FIG. 4b





6/13

FIG. 5a

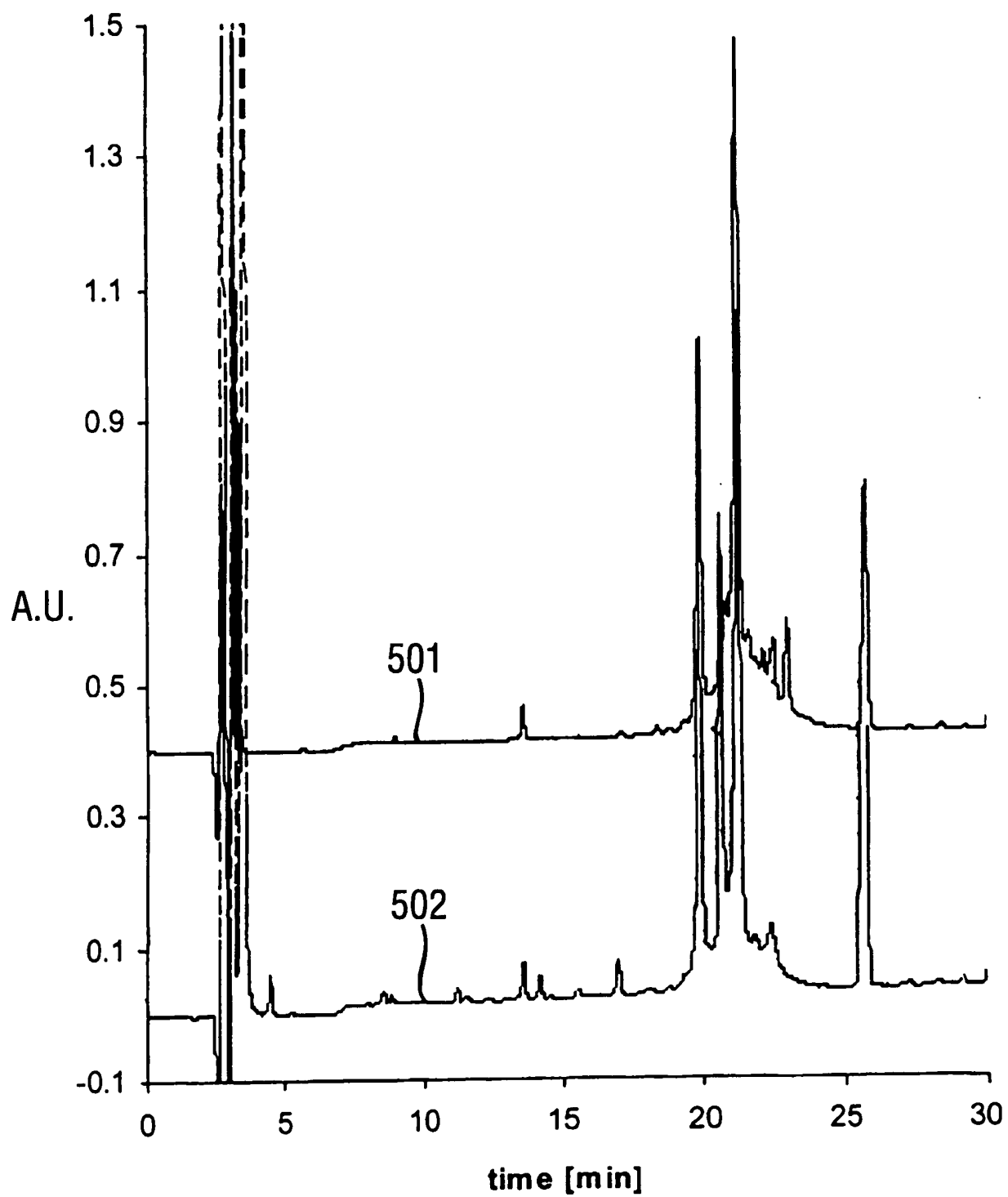
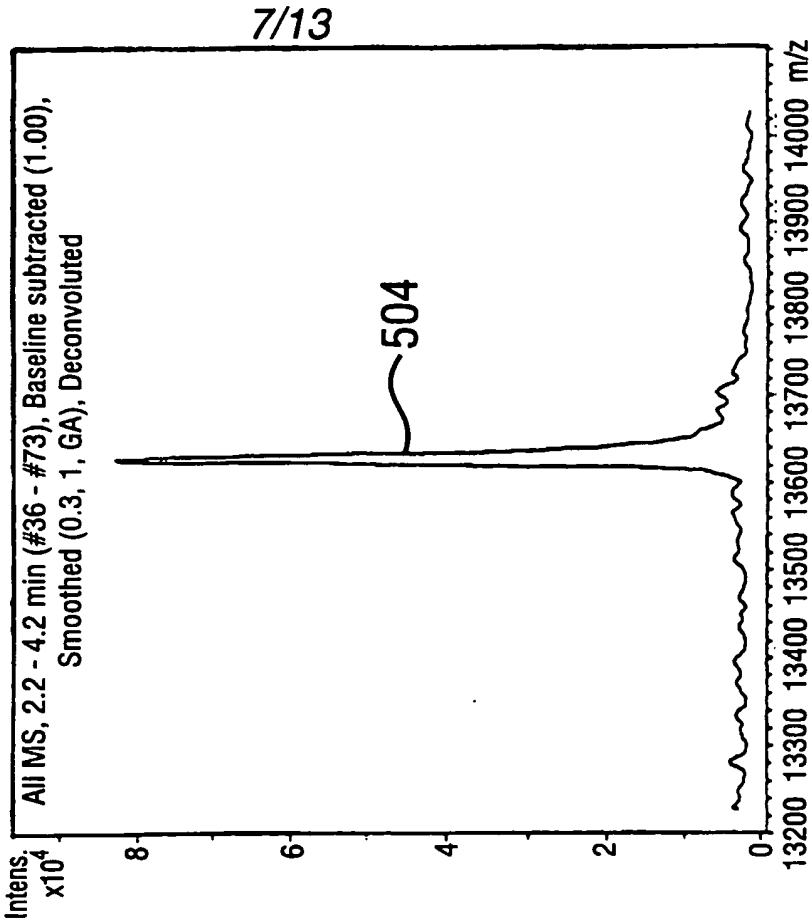
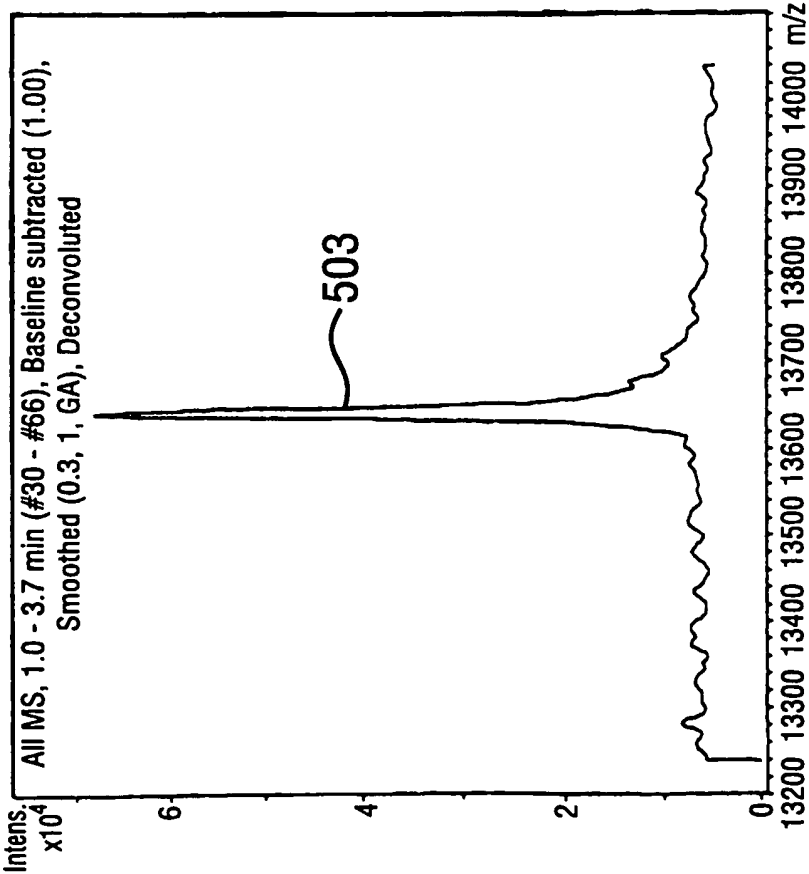
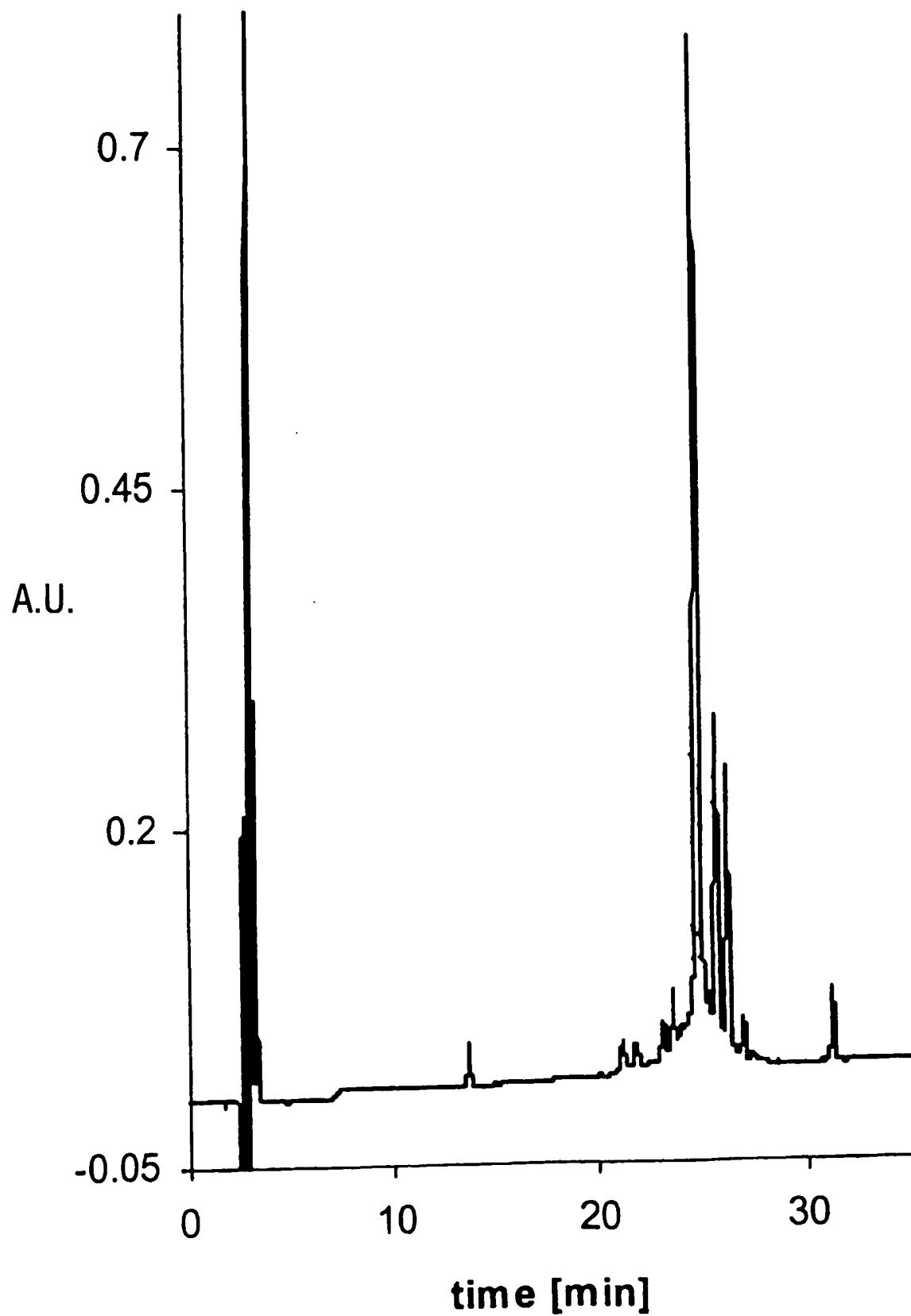


FIG. 5b



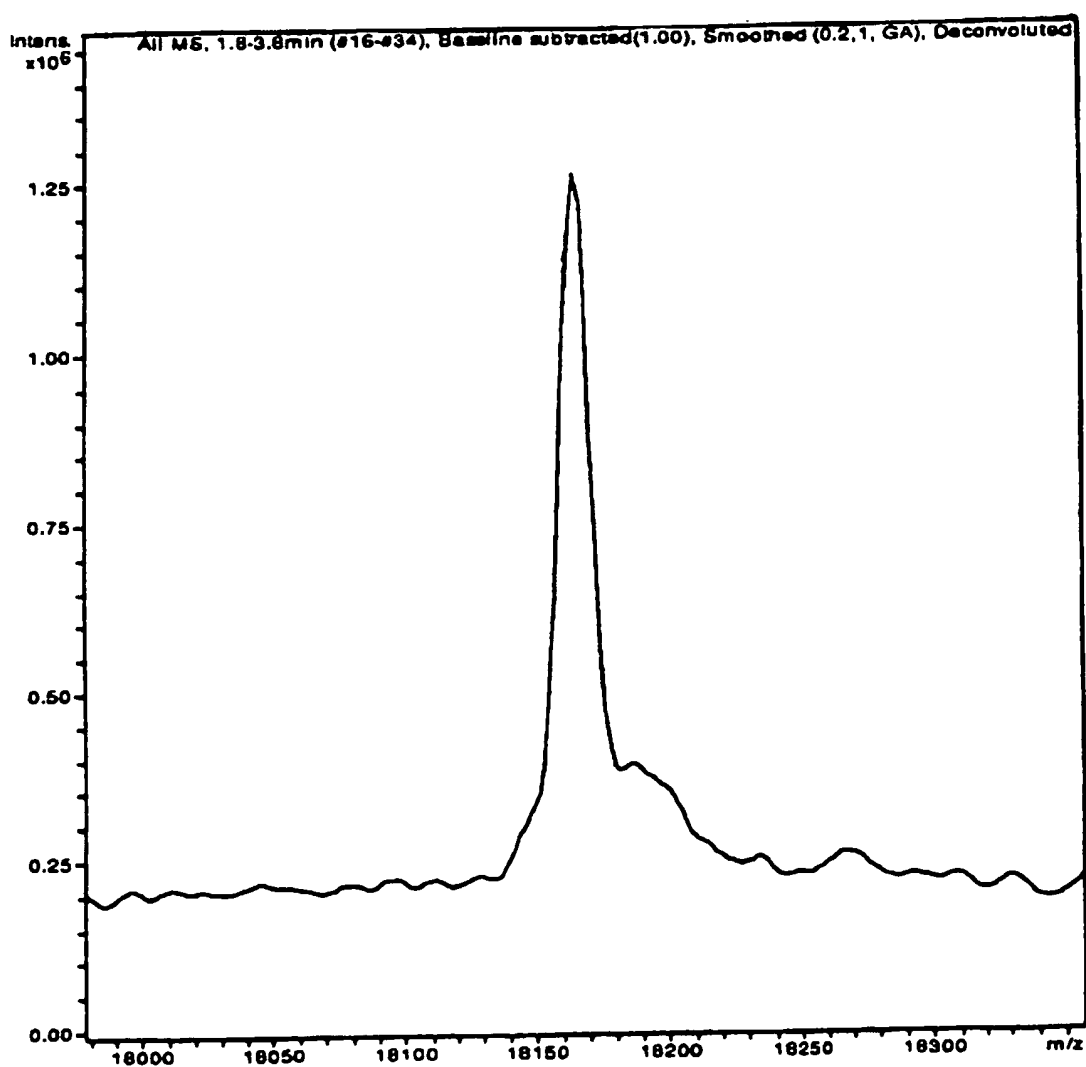
8/13

FIG. 6a



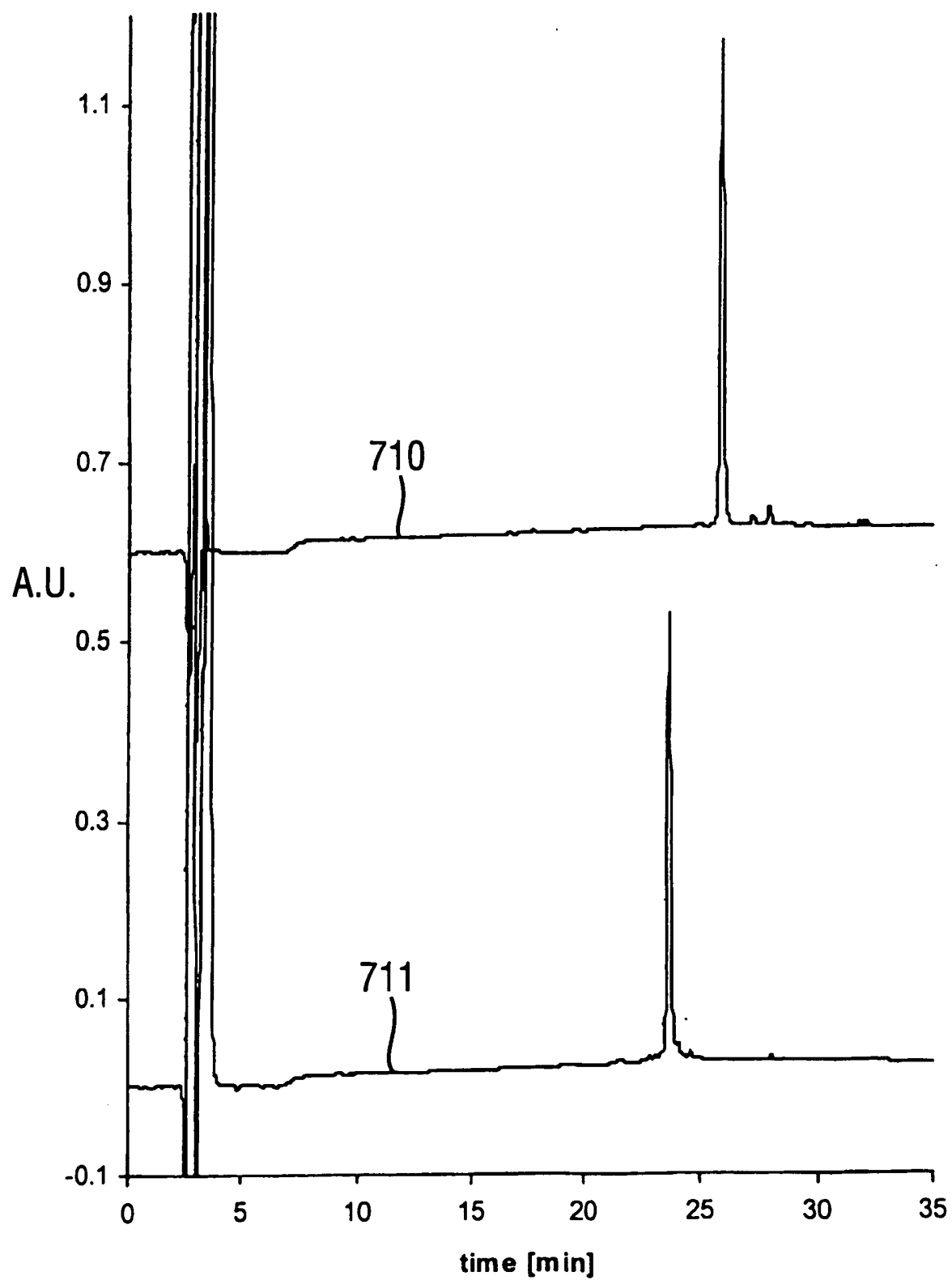
9/13

FIG. 6b



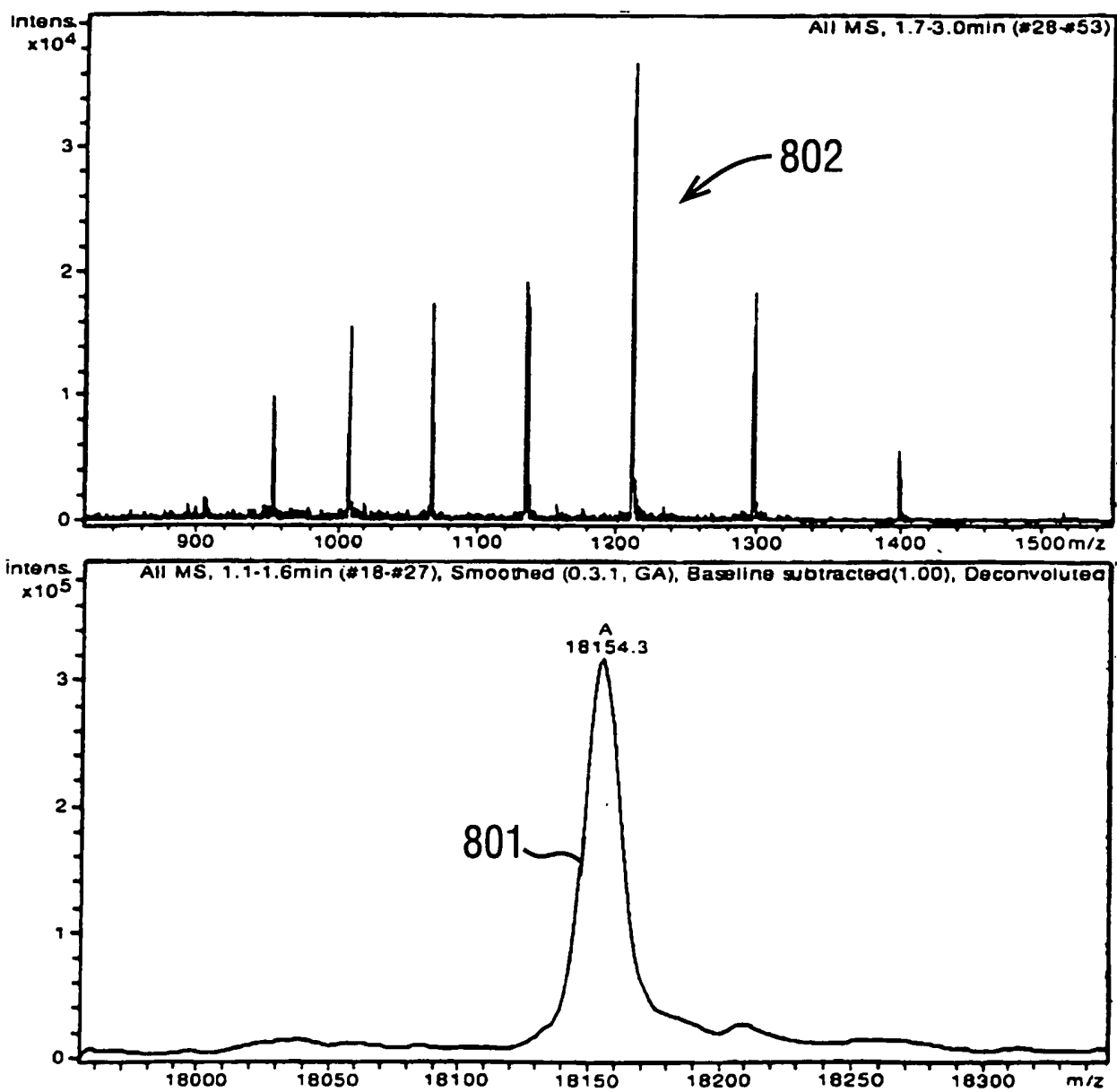
10/13

FIG. 7



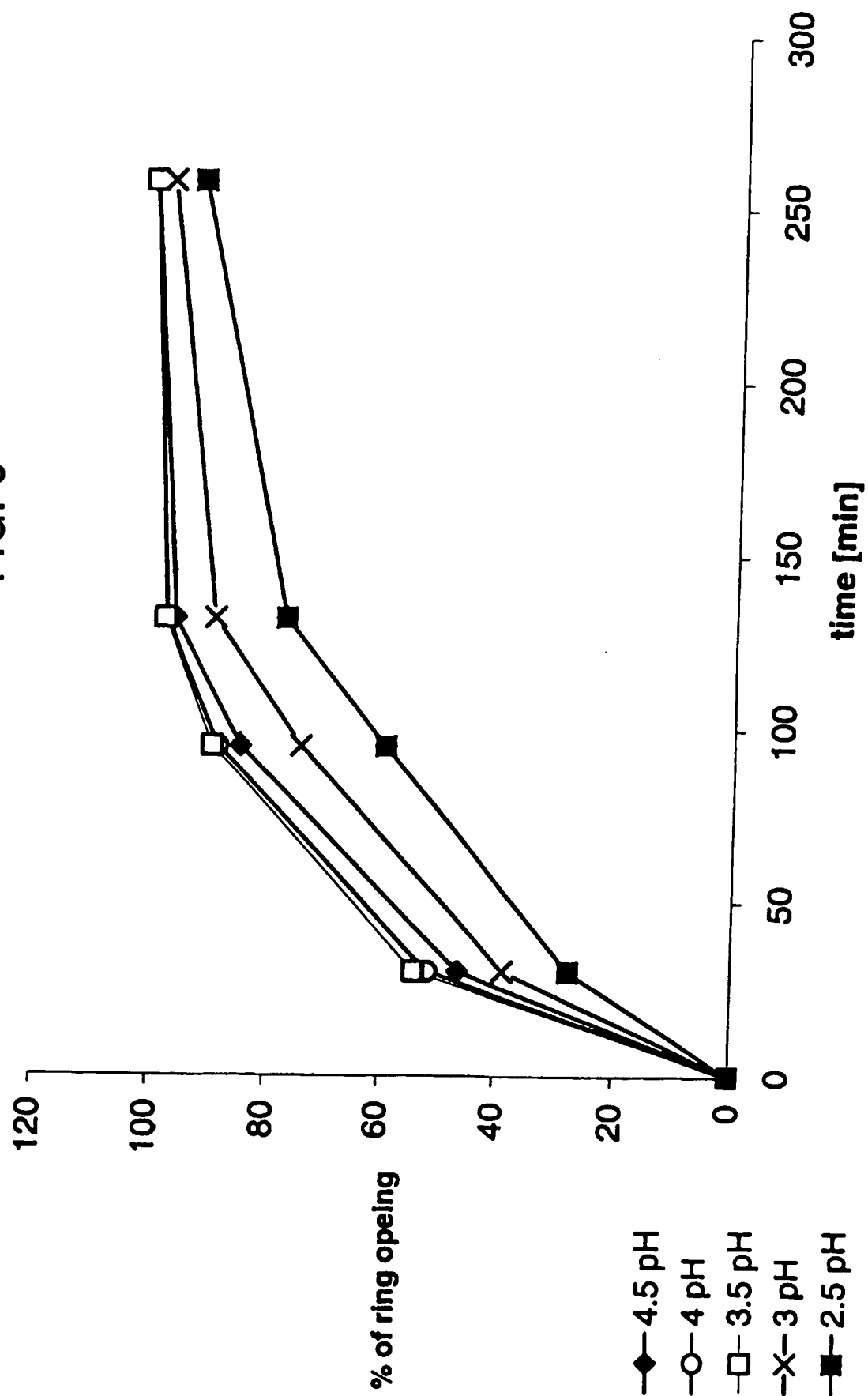
11/13

FIG. 8



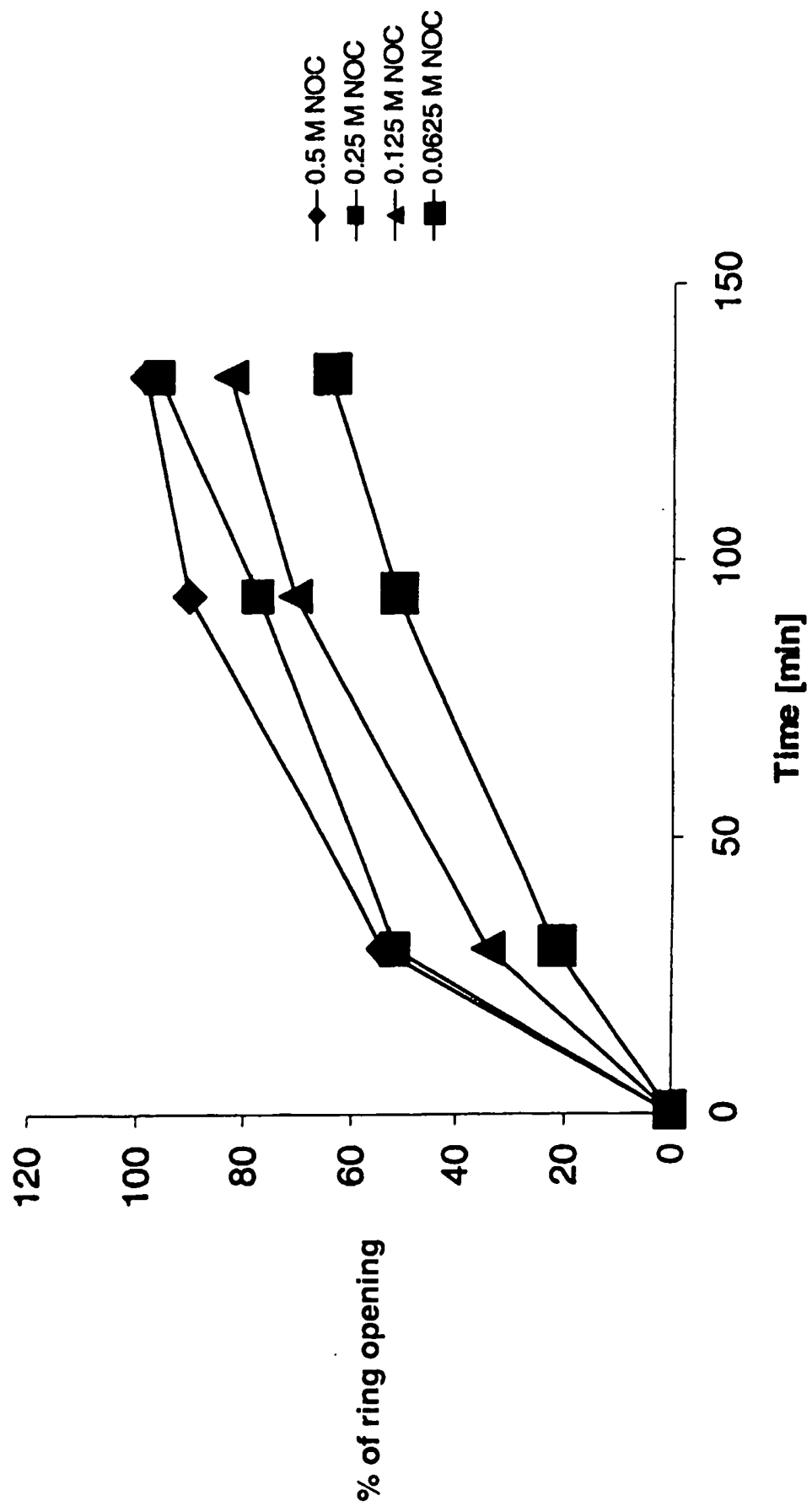
12/13

FIG. 9



13/13

FIG. 10





## SEQUENCE LISTING

```
<110> Villain, Matteo  
Gaertner, Hubert  
<120> Improved Native Chemical Ligation with Three or More Components  
<130> 5002  
<140>  
<141>  
<160> 7  
<170> Microsoft Word 2000
```

```
<210> 1
<211> 29
<212> PRT
<213> Homo Sapiens
<400> 1
```

Cys Ala Glu Lys Ser Asp Tyr Ile Arg Lys Ile Asn Glu Leu Met Pro  
5 10 15

Lys Tyr Ala Pro Lys Ala Ala Ser Ala Arg Thr Asp Leu  
20 25

```

<210>      2
<211>      48
<212>      PRT
<213>      Homo Sapiens
<220>
<221>      Modified amino acid
<222>      (1)
<223>      Thioproline
<220>
<221>      Modified amino acid
<222>      (48)
<223>      Glycine thiophenylester
<400>      2

```

Xaa Glu Lys Leu Lys Lys Leu Asp Ser Gln Ile Cys Glu Leu Lys Tyr  
5 10 15

Asp Lys Gln Ile Asp Leu Ser Thr Val Asp Leu Lys Lys Leu Arg Val  
20 25 30

Lys Glu Leu Cys Lys Ile Leu Asp Asp Trp Gly Glu Thr Cys Lys Gly  
35 40 45

```

<210> 3
<211> 42
<212> PRT
<213> Homo Sapiens
<220>
<221> Modified amino acid
<222> (1)
<223> Thioproline
<220>
<221> Modified amino acid
<222> (42)
<223> Isoleucine thiophenylester
<400> 3

```

Xaa Arg Glu Ala Arg Gly Lys Glu Asn Arg Leu Cys Tyr Tyr Ile Gly  
5 10 15

Ala Thr Asp Asp Ala Ala Thr Lys Ile Ile Asn Gly Val Ser Lys Pro  
20 25 30

Leu Ala His His Ile Pro Val Gly Lys Ile  
35 40

```
<210> 4
<211> 39
<212> PRT
<213> Homo Sapiens
<220>
<221> Modified amino acid
<222> (39)
<223> phenylalanine thiophenylester
<400> 4
```

Leu Arg Pro Gly Asp Cys Glu Val Cys Ile Ser Tyr Leu Gly Arg Phe  
5 10 15

Tyr Gln Asp Leu Lys Asp Arg Asp Val Thr Phe Ser Pro Ala Thr Ile  
20 25 30

Glu Asn Glu Leu Ile Lys Phe  
35

```
<210> 5
<211> 158
<212> PRT
<213> Homo Sapiens
<400> 5
```

Leu	Arg	Pro	Gly	Asp	Cys	Glu	Val	Cys	Ile	Ser	Tyr	Leu	Gly	Arg	Phe
				5					10					15	
Tyr	Gln	Asp	Leu	Lys	Asp	Arg	Asp	Val	Thr	Phe	Ser	Pro	Ala	Thr	Ile
			20					25					30		
Glu	Asn	Glu	Leu	Ile	Lys	Phe	Cys	Arg	Glu	Ala	Arg	Gly	Lys	Glu	Asn
		35					40					45			
Arg	Leu	Cys	Tyr	Tyr	Ile	Gly	Ala	Thr	Asp	Asp	Ala	Ala	Thr	Lys	Ile
	50					55					60				
Ile	Asn	Gly	Val	Ser	Lys	Pro	Leu	Ala	His	His	Ile	Pro	Val	Gly	Lys
65					70					75					80
Ile	Cys	Glu	Lys	Leu	Lys	Lys	Leu	Asp	Ser	Gln	Ile	Cys	Glu	Leu	Lys
				85					90					95	
Tyr	Asp	Lys	Gln	Ile	Asp	Leu	Ser	Thr	Val	Asp	Leu	Lys	Lys	Leu	Arg
			100					105					110		
Val	Lys	Glu	Leu	Cys	Lys	Ile	Leu	Asp	Asp	Trp	Gly	Glu	Thr	Cys	Lys
		115					120					125			
Gly	Cys	Ala	Glu	Lys	Ser	Asp	Tyr	Ile	Arg	Lys	Ile	Asn	Glu	Leu	Met
	130					135					140				
Pro	Lys	Tyr	Ala	Pro	Lys	Ala	Ala	Ser	Ala	Arg	Thr	Asp	Leu		
145					150					155					

```
<210> 6
<211> 8
<212> PRT
<213> Homo Sapiens
<400> 6
```

Cys Tyr Ala Lys Tyr Ala Lys Leu  
5

```
<210> 7
<211> 8
<212> PRT
<213> Homo Sapiens
```

<220>  
<221> Modified amino acid  
<222> (1)  
<223> thioproline

<400> 7

Xaa Tyr Ala Lys Tyr Ala Lys Leu  
5